DOCUMENT RESUME

ED 351 410 UD 028 915

AUTHOR Sem, Wee Cheng
TITLE Report on Student Performance in the Northern Territory Primary Assessment Program for Urban Schools: 1986.
INSTITUTION Northern Territory Dept. of Education, Darwin (Australia).
REPORT NO ISBN-0-7245-1335-3
PUB DATE May 87
NOTE 219p.
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS *Achievement Gains; Achievement Tests; Comparative Testing; Educational Assessment; *Elementary School Students; Federal Programs; Foreign Countries; Grade 5; Grade 7; Intermediate Grades; Junior High Schools; Junior High School Students; *Mathematics Achievement; Program Evaluation; *Reading Achievement; Standardized Tests; *Test Results; *Urban Schools
IDENTIFIERS *Australia (Northern Territory)

ABSTRACT

This publication reports on student achievement in the basic skills of reading and mathematics at Years 5 and 7 in 52 urban schools of the Northern Territory of Australia in 1986, the third year of implementation of the Primary Assessment Program (PAP). PAP has been acknowledged by the national committee of the Australian Cooperative Assessment Program as significant because of its role in providing quality evaluation and assessment strategies for use by schools and teachers to review and improve teaching and learning. The report consists of 13 chapters, most of which offer detailed analyses of various sections of the assessment tests, including examples from the testing instruments. Early chapters of the report discuss the criterion referenced testing used in the program and the program's approach to data collection and analysis. An evaluation of the results indicates that for the majority of questions in the reading test and mathematics test, the success rates were generally satisfactory. Performance was particularly good in routine computation and measurement in mathematics. Areas of weakness are reported to be decimal computation in Year 7 and comprehension of prose passages in Year 5. Included are numerous sample test items, 14 tables, an appendix containing 39 additional figures illustrating results, and 14 references. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED351410

REPORT ON STUDENT PERFORMANCE IN THE NORTHERN TERRITORY PRIMARY ASSESSMENT PROGRAM FOR URBAN SCHOOLS: 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. Fowler
N. Ter. Dept of Education
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED028915

NORTHERN TERRITORY DEPARTMENT OF EDUCATION

BEST COPY AVAILABLE

REPORT ON STUDENT PERFORMANCE IN THE NORTHERN TERRITORY PRIMARY ASSESSMENT PROGRAM FOR URBAN SCHOOLS: 1986

WEE CHENG SEM
ACCREDITATION & ASSESSMENT SECTION
CURRICULUM & ASSESMENT BRANCH
NT DEPARTMENT OF EDUCATION

Northern Territory Department of Education
Darwin May 1987

ISBN 0 7245 1335 3

Printed by the Distance Education Branch
of the Northern Territory Department of Education
May 1987
P&P87/314-200

FOREWORD

The Primary Assessment Program has been in operation for three years now in urban schools throughout the Northern Territory. It has been acknowledged by the national committee of ACAP (the Australian Cooperative Assessment Program) as a significant project because of its role in 'providing quality evaluation and assessment strategies for use by schools and teachers to review and improve teaching and learning'. From the beginning, the main thrust of this program has been to assist and improve instruction in the core curriculum in reading and mathematics.

A second important feature of this program is monitoring standards of achievement in the basic skills of reading and mathematics. This third report on student performance in core knowledge and skills at Years 5 and 7 in urban schools is based on the results of a testing program carried out in 1986.

A close examination of the results indicates that for the majority of questions in the reading and mathematics test, the success rates were generally satisfactory. Performance was particularly pleasing in areas such as routine computation and measurement in mathematics. It appears also that most students responded well to situations in which basic literacy skills were required as the results in reading for different purposes indicated.

The report has highlighted some areas of concern such as decimal computation in Year 7 and comprehension of prose passages in Year 5. These areas of weakness identified in the report will be examined and remedial action taken at both the school and system levels. The utilisation of the Primary Assessment Program in improving standards and charting the path to excellence in NT education will be a major consideration over the next few months.

I wish to thank all principals and staff for their continued support and cooperation in the administration of the primary assessment program. Officers of Curriculum and Assessment Branch have made special efforts to produce this valuable report. The report confirms the view that the primary assessment program is a dynamic assessment tool with the capacity to improve and expand to meet changing needs within our schools and to serve as a stimulus to the improvement of Territory education.

G J SPRING
Secretary

TABLE OF CONTENTS

		Page
Foreword		(iii)
Table of Contents		(iv)
List of Tables		(v)
Members of Primary Assessment Committee		(vi)
Primary Assessment Committee : Terms of Reference		(vii)
Acknowledgements		(viii)
Chapter One	Criterion-referenced Testing: Its Application in the Primary Assessment Program	1
Chapter Two	Data Collection	8
Chapter Three	Data Analysis	12
Chapter Four	Analysis of Performance on the Mathematics Tests : Phase 3	17
Chapter Five	Comparison between 1985 and 1986 Performance : Mathematics Phase 3	28
Chapter Six	Results on Selected Tests and Items: Mathematics Phase 3	33
Chapter Seven	Analysis of Performance on the Mathematics Tests: Phase 2	50
Chapter Eight	Comparison between 1985 and 1986 Performance : Mathematics Phase 2	60
Chapter Nine	Results on Selected Tests and Items : Mathematics Phase 2	64
Chapter Ten	Analysis of Performance on the Reading Tests : Stage 7	83
Chapter Eleven	Analysis of Performance on the Reading Tests : Stage 5	96
Chapter Twelve	Comparison between 1985 and 1986 Performance : Reading Stages 5 and 7	108
Chapter Thirteen	Summary and Conclusion	115
Appendix		130
References		169

LIST OF TABLES

		Page
TABLES		
1A	Percentages at Different Cutoffs: Mathematics Phase 3	21
1B	Percentages Achieving Competence: Mathematics Phase 3	24
1C	Changes in Performance: Mathematics Phase 3	31
2A	Percentages at Different Cutoffs: Mathematics Phase 2	54
2B	Percentages Achieving Competence: Mathematics Phase 2	57
2C	Changes in Performance: Mathematics Phase 2	62
3A	Percentages at Different Cutoffs: Reading Stage 7	91
3B	Percentages Achieving Competence: Reading Stage 7	93
4A	Percentages at Different Cutoffs: Reading Stage 5	104
4B	Percentages Achieving Competence: Reading Stage 5	106
3C	Changes in Performance: Reading Stage 7	111
4C	Changes in Performance: Reading Stage 5	112
3D	Reading Tests: Year 7 Changes in Performance: Summary	113
4D	Reading Tests: Year 5 Changes in Performance: Summary	114

NT BOARD OF STUDIES

PRIMARY ASSESSMENT COMMITTEE

Dr Harry Payne	Chairman Director, Curriculum and Assessment Branch
Mr John Pollock	A/Director, Schools North
Mr Jon Cook	A/Superintendent Curriculum (Observer)
Mr Rowan Barnsley	Principal Education Officer Accreditation and Assessment
Mr Wee Cheng Sem	Senior Research Officer Accreditation and Assessment
Mr Ray Hopkins	A/Superintendent Primary South
Mr Ron Abbott	Principal, Millner Primary School
Ms Barbara Littler	A/Executive Officer NT Board of Studies

PRIMARY ASSESSMENT COMMITTEE

TERMS OF REFERENCE

(1) Make recommendations to the Board on policy for the assessment of student competence in core skills and understanding at Years 5 and 7 levels in urban primary and Aboriginal community schools.

(2) Make recommendations to the Board on policy for the screening and assessment of students in the early childhood years.

(3) Provide guidelines to relevant committees on the development of instruments and procedures for the assessment of student competence in core skills and understandings.

(4) Oversight the work of committees as indicated in (3) above.

(5) Report to the Board on areas of strengths and weakness as identified by assessment programs oversighted by the Board and make recommendations to the Board concerning appropriate actions which might arise therefrom.

(6) Oversight the development and use of assessment materials beyond the core but within the recommended curriculum, as and when required by the Department.

ACKNOWLEDGEMENTS

The assistance and services of the following people are acknowledged with grateful thanks:

Principals and staff of urban primary schools in the Northern Territory for administering the tests and returning the students' answer sheets.

Dr John Izard, Assistant Director, Australian Council for Educational Research, who was a consultant during the earlier phase of the project.

Ms Glenda Cunningham for typing the report and making all the necessary changes to the draft.

Ms Stoula Poullas and **Ms Cherie Murray** for keying in the data from the students' answer sheets and checking for errors.

Staff at the Information Centre, NCOM, for assistance in data processing and programming.

CHAPTER ONE

CRITERION-REFERENCED TESTING: ITS APPLICATION IN THE PRIMARY ASSESSMENT PROGRAM

ABSTRACT

The distinction between norm-referenced tests (NRTs) and criterion-referenced tests (CRTs) is pointed out. A competency test is defined as a particular kind of CRT. The amplified objectives approach, a procedure for generating test items, is described. Ways for improving item quality are outlined. Finally, the issue of performance standards in CRT is introduced.

1.1 INTRODUCTION

This brief introduction to criterion-referenced measurement is intended to give the reader only an appreciation of the subject. For the more interested reader, the bibliography provides a select list of references.

In criterion-referenced testing, a student's performance is judged against a defined criterion. The criterion, as used in this program, is the mastery or competency level. For instance, given a set of 10 questions involving the ability to recall basic facts on tables (up to 12), can the student get 9 out of 10 correct? The competency level, used in this example, is 9 out of 10 correct. The score a student gets in this test will indicate whether he or she is competent on tables.

In a typical classroom situation, the teacher grades a student on a test using percentage or letter grading. For instance, Lyn's score on a Mathematics test was 96% or an 'A'. Quite often, this widely used marking system does not fit the definition of criterion-referenced measurement because of the following reasons:

1. The % score or grade does not describe what Lyn can or cannot do in Mathematics. It does not indicate what knowledge or skill is involved; it is simply a score on a Mathematics test. The % score, however, does allow one student to be compared against another on the same test.

2. It is probable that the test was built without the benefit of a prior statement of behavioural objectives for the instructional unit.

3. Even when objectives are clearly defined, the use of a single score to indicate performance on a number of objectives is not very helpful for instructional purposes. The use of a single score can easily mask what a student can actually do.

It is useful to note the distinction between NRTs and CRTs. NRTs yield scores with the overriding emphasis on comparing one student against another. To achieve this goal, the test developer selects items that will 'spread out' the scores. Items of medium difficulty will do that job pretty well. On the other hand, CRTs yield scores with the overriding emphasis on giving a clearer picture of what each student's score means regardless of how that score compares with the scores of other students. Each score is referenced to a specific objective. For example, a score of 9 out of 10 in measuring angles between 0 and 180 degrees.

1.2 DEFINITION OF A COMPETENCY TEST

A competency test is simply a particular kind of CRT and therefore, like a CRT, it must be developed and used in ways different from a NRT.

The tests developed for the Northern Territory Primary Assessment Program are competency tests. They are designed to determine a student's level of performance in relation to the objective being measured. A standard is set at 80% correct, for instance. Each student's test score is interpreted relative to this criterion level.

1.3 AMPLIFIED OBJECTIVES APPROACH

Of all the models employed to generate CR test items, the amplified objectives model advocated by Popham (Popham, 1978) is perhaps the most popular and widely used. This approach begins with the objective. Additional information is then added to reduce as much as possible the ambiguities typically associated with the content and the testing situation. The approach adopted to create test items for the Primary Assessment Program was similar. It began with the core objective. The objective was then expanded to describe clearly the limits for the item content.

EXAMPLE Multiply 3-digit whole numbers by single digit whole numbers.

One version of the amplified objectives approach consists of:

1. a response description, which is an instructional objective of interest that also includes information on the type of response that is required;

2. content limits, which is essentially a rule for determining the content relevant to the achievement of the objective;

3. an item format specification, which gives a detailed description of the characteristics of the items; and

4. item directions and instructions, which tell the examinee how to respond to the item.

EXAMPLE (POPHAM, 1978)

1. Response description: when presented with sketches of various clock faces showing different times, the student will be able to select the correct time from numerical alternatives.

2. Content limits: the clock faces should be round or square and the time shown in five minute units. The minute and hour hands should be shown by long and short (respectively) arrows or straight lines.

3. Item format: four answer options will be given, one of which is the correct answer.

Incorrect answers may be:

the time that would be depicted if the minute and hour hand had been reversed

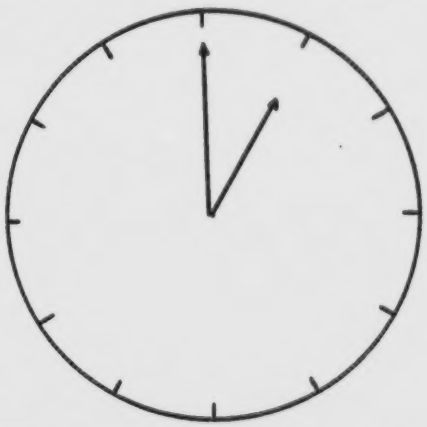
a time 5 min earlier or later than the correct time

the transposal of the correct time, e.g. 30:8 instead of 8:30

a time randomly selected

SAMPLE ITEM

For the clock pictured on the left, circle A, B, C or D that shows the correct time.



A 12:05

B 1:00

C 1:05

D 12:00

1.4 PROCEDURES FOR IMPROVING ITEM QUALITY

There are several procedures which can be used to identify items that are 'faulty'. In order for the test to have content validity, the items must fit the objective being measured. Unless one can say with confidence that the items in a competency test measure the intended objective, any interpretation of the test scores is questionable. In the Primary Assessment Program, this was done by inspection.

FACILITY

In selecting items for inclusion in the tests, several characteristics possessed by items were examined. Among these were the item's facility and discrimination.

The facility of an item was a very important characteristic used in selecting items for the program. This is indicated by the proportion of students who got the item right.

EXAMPLE

If in a class of 20 students, the number who got an item right was 12, the facility of the item is given by:

$$\text{Facility} = \frac{\text{Number who got the item right}}{\text{Total number who attempted}}$$

$$= \frac{12}{20}$$

$$= 0.60$$

Facility can have a range of values from 1.00 where everyone got the item right to 0.00 where no one got the item right.

The facility or p-value was a useful indicator for analysis. In the CR test situation, high p-values could be expected if what was measured by the items had been taught well. A high or a low p-value did not mean a bad item. If the p-value was high or very low, the item was checked to see if it discriminated between mastery and non-mastery students.

A good way to use the facility value in a class situation is to obtain one before instruction and another after instruction. If, however, the p-values were close, one may examine the instruction or the test items.

DISCRIMINATION

Another characteristic used in the analysis was the item's discrimination index. Since the aim of a CRT is to tell the teacher whether the student is competent or not on a particular objective, it was important to select items that discriminated between students who were competent (masters) and those who were not (non-masters).

An item discriminates if a higher proportion in the competent group (mastery group) got the item right compared with the proportion in the non-competent group (non-mastery group) who also got the item right. The competent group comprises students who have obtained 80% of the number of items correct where 80% has been defined as the competency level.

The discrimination index is given by:

$$\text{Discrimination Index} = \frac{\text{No of students in mastery group who got item right}}{\text{Total no of students in mastery group}} - \frac{\text{No of students in non-mastery group who got item right}}{\text{Total no of students in non-mastery group}}$$

EXAMPLE

In a class of 20 students, 14 students achieved the mastery level of 80% correct on a test. This means that

No of students in mastery group	=	14
No of students in non-mastery group	=	6

Suppose, on a particular item in the test the number of students in the mastery group who got the item right was 12, and the number in the non-mastery group who also got the item right was 2.

Then, the discrimination index is given by:

$$\text{Discrimination Index} = \frac{12}{14} - \frac{2}{6}$$
$$= 0.86 - 0.33$$
$$= 0.53$$

A discrimination index of 0.53 indicates a highly discriminating item. A low value such as 0.10 or less, or a negative value, is a warning flag.

When a seemingly poor item was located, one had to decide whether to alter or discard it. Other characteristics of CRTs were also examined, but these would take too much space to explain here. The panels involved in the development of the item pools for this program selected items that had certain characteristics that 'good' CRTs possess.

1.5 STANDARDS OF PERFORMANCE

One of the most perplexing problems facing teachers involves the necessity of deciding upon the level of student achievement that constitutes an acceptable performance. Several methods have been reported in the measurement literature on setting performance standards (Hambleton, et al 1978).

However, one clear fact that has emerged from all these standard-setting procedures is this: all standard-setting methods are arbitrary. They are arbitrary because they involve judgements of one kind or another.

In the Northern Territory Primary Assessment Program, for instance, the standard employed for acceptable performance is 80% correct; that is, a student who answers correctly 80% of a set of items measuring a particular objective is considered to be competent on that objective. This performance standard applicable to most tests in the program was established by the panels of item writers. However, the 80% correct level is not applied universally to all the tests in the item pools. Some tests have lower cutoff scores; others have higher (Wee, 1984).

The panels considered these factors when setting standards: the importance of the objective, its relationship to other objectives, the number of items in the test and the nature of the test.

CHAPTER TWO

DATA COLLECTION

ABSTRACT

1986 was the third year of implementation of the Primary Assessment Program for all urban schools. There were fifty-one schools in the target population. The schools had been assigned into four groups, each group consisting of large, medium and small schools (based on total enrolment) from all regions, as far as possible. Students' answer sheets were returned from each school after the tests had been administered. Confidentiality of information was protected as each school was identified by a code and students were identified by numbers instead of names. Students tested were in Years 5 and 7. As in previous years, the cooperation of principals and staff involved in the testing was indeed extremely good.

2.1 SCHOOLS INCLUDED IN THE PRIMARY ASSESSMENT PROGRAM

During 1986, urban primary schools throughout the Northern Territory participated in the Primary Student Assessment Program designed to determine the levels of student achievement in the core areas of Reading and Mathematics at Years 5 and 7. It was the third year of implementation of this program.

All urban schools, with the exception of the two schools of the air, were included in the target population for the collection of data. Fifty-one schools were identified in Darwin, Alice Springs, Katherine, Tennant Creek, Jabiru, Nhulunbuy, Alyangula and Batchelor.

The schools had been assigned into four groups. Each group included large, medium and small schools (based on total student enrolment) and, as far as possible, schools from each region. Sampling was done by region and school size.

2.2 STUDENTS INCLUDED IN THE PROGRAM

Students in Years 5 and 7 were the target population. The following definitions of Year 5 and Year 7 were used:

YEAR 5:

'Children who are in their 5+ year at primary school and are currently undertaking study of the core skills and understandings at or above Phase Two in Mathematics and Stage Five in English. This means that not all children will be tested on all objectives at Year 5'.

YEAR 7:

'Children who are judged by the school to be in their final year at primary school. This also means that all children in Year 7 will be tested'.

'In the case of ungraded classes, students who should be in the target population may be determined on the basis of their ages'.

2.3 CONFIDENTIALITY OF INFORMATION

To protect the confidentiality of information collected, each school was given a school code known only to the project coordinator. Each student in Year 5 and Year 7 was also given a student number by the school. The student used the same number for all the tests that he or she took. The school code and student number were used to identify each answer sheet keyed into the computer for data processing and analysis.

2.4 SAMPLING OF TESTS

Each school was asked to administer a sample of tests in Reading and Mathematics at each of the two year levels. The tests were selected from the item pools, and then randomly assigned to each of the four groups of schools. One of the reasons for test sampling was that each school in a particular group was required to administer only a fraction of the tests from the item pool.

The distribution of tests to the groups was as follows:

	NUMBER OF TESTS			
GROUP	MATHS/P3	MATHS P/2	READING S/7	READING/S5
1	6	6	5	5
2	6	6	5	5
3	6	6	5	5
4	5	6	5	5

The tests varied in test length, i.e. the number of items in each test. In terms of the total number of items in each subject area and the relative difficulties of the tests within each subject area, the distribution was found to be fairly even.

2.5 STUDENTS' ANSWER SHEETS

The following information was supplied by each school on the students' answer sheets:

School code
Student number
Test code
Date (i.e. the date the test was administered)

2.6 BLANK ANSWER SHEETS

The tests were administered to students in Years 5 and 7 with the exception of a very small number in Year 5 who, in the opinion of the teachers, were not able to cope with the tests. These were students who would have guessed at random the answer to every test item without having understood the concept.

As all students in Year 7 were to be tested, it was necessary to know how many did not do any of the tests because the questions were well beyond their ability to handle. Teachers used their own discretion in this matter as they knew which students would tend to make random guesses at the answers to practically all the items. For such students, a blank answer sheet with the student number corresponding to the test was requested.

The number of blank answer sheets returned was indeed very few although no count was made of them. However, if schools submitted blank answer sheets with student numbers, these were entered together with the rest. Blank answer sheets were given zero marks.

2.7 ADMINISTRATION OF THE TEST

No standard administration procedures were provided to teachers as the administration of these tests was seen as part of the normal instructional process. No time limits were specified: teachers were asked to use their own judgement in deciding whether enough time had been given.

The timing and sequence of testing were also to be decided by the classroom teacher.

2.8 TIME FOR RETURN OF ANSWER SHEETS

The students' answer sheets were sent in as they became available during the first and second semester of 1986. In actual fact, most schools returned their students' answer sheets during October and November. There was a very high level of cooperation on the part of principals and school staff.

CHAPTER THREE

DATA ANALYSIS

ABSTRACT

The results have been analysed and reported in subsequent chapters as follows:

1. A set of tables shows percentages of students responding correctly to 70%, 80%, 90% and 100% of items for the Mathematics tests. In the case of Reading, the 60% cutoff has also been included as a reference point.

2. Another set of tables shows percentages of students achieving competence on each test/objectives.

3. Graphs are provided showing percentages of students achieving competence on each test. Estimates of sampling error are included in the graphs. These graphs may serve as a guide for comparing individual school results against the Territory results.

4. A diagram illustrates overall performance in each subject area for each year level.

5. Comparisons are made between the results for 1985 and 1986 on each test and changes in performance are discussed. The time of testing was approximately the same for both years.

3.1. METHOD OF REPORTING

As a means of providing more information to schools, percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items within a test have been reported for the Mathematics tests. In the case of Reading, a further reference point, i.e. the 60% cutoff, has been included.

TABLES

Table 1A (Chapter 4) and Table 2A (Chapter 7) show percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items for the Mathematics tests in Phase 3 (Year 7) and Phase 2 (Year 5) respectively.

Table 3A (Chapter 10) and Table 4A (Chapter 11) show percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items for the Reading tests in Stage 7 (Year 7) and Stage 5 (Year 5) respectively.

The next set of tables within each of the chapters shows percentages of students achieving competence on each test. Table 1B (Chapter 4) and Table 2B (Chapter 7) show percentages of students achieving competence on each of the Mathematics tests in Phase 3 and Phase 2 respectively. The competency level was established at 80% correct.

Similarly, Table 3B (Chapter 10) and Table 4B (Chapter 11) show percentages of students achieving competence on each of the Reading tests in Stage 7 and Stage 5 respectively, where the competency level was 70% correct.

Further tables are included in Chapters 5, 9 and 12 to show changes in performance between 1985 and 1986.

GRAPHS

Graphs showing percentages of students achieving competence on each test in Mathematics and Reading are illustrated in the Appendix. The graphs include sampling errors for proportions to enable schools to compare the results of their own students with those of the Territory.

3.2 SAMPLING ERROR FOR PROPORTIONS

For simple random samples, the standard error for proportions is estimated by calculating:

$$Sp = \sqrt{\frac{p(1-p)}{n-1}}$$

Kish (1965)

where p = sample proportion achieving the cutoff score
n = sample size

However, it has been shown that the use of this formula based on simple random sampling seriously underestimates the magnitude of sampling error when intact groups in schools are used (Ross, 1976). In this report, estimates of sampling error were adjusted in accordance with recommendations by Ross.

3.3 COMPARING SCHOOL RESULTS WITH NORTHERN TERRITORY RESULTS

To enable schools to compare the results of their own students with those of the Territory, the following adjustments have been made to the calculations based on n = 300, for the 95 per cent confidence limits (Kish, 1965):

Value of p	Sampling error (95 per cent confidence limits)
0.50	8% points
0.60	8% points
0.70	7% points
0.80	6% points
0.90	5% points
0.95	4% points

(Note: p = proportion achieving competence on each test)

By determining confidence limits of 95 per cent for a particular proportion, p, based on sample data, an interval could be established, so that in 95 per cent of similar computations, the interval would contain the population proportion.

For instance, if the sample proportion achieving competence on a particular test is 90% (p = 0.90), the 95 per cent confidence limits are $90 \pm 5\%$, i.e. 85% and 95% from the above calculations. The interval is between 85% and 95%. That is, between 85% and 95% of the population achieve competence on the test. This is a more likely estimate than 90%.

The sampling error is only an estimate but it may serve as a useful guide to schools when comparing the performance of their own students against the performance of Territory students on the same test.

To determine whether differences in the proportions of students achieving competence are significant, the estimated sampling error for proportions given above should be used as a guide.

The graphs in the Appendix incorporate the sampling errors. Individual school results on each test may be compared against the Territory results simply by reading from the graph.

EXAMPLE

Let us say that for a particular test, 68.0% of Territory students achieved competence. This approximates to 0.70 for the value of p.

The sampling error estimated for p = 0.70 (or 70%) achieving competence is 7% points. This means that results obtained by a school would need to differ by 7% points to be statistically significant.

In other words, if the results obtained by a school show that more than 75% (68+7) of its students have achieved competence in this particular test, the school results would appear to be significantly better than the Territory results. On the other hand, if the results obtained show that less than 61% (68-7) of its students have achieved competence in this test, then the school results would seem to be significantly worse than the Territory results.

The sampling errors estimated are also shown in Tables 1A, 2A, 3A and 4A in the chapters that follow.

Graphs have been drawn that incorporate these sampling errors. Simply by reading from the graph, individual school results on each test may be compared against the Territory results.

CHAPTER FOUR

ANALYSIS OF PERFORMANCE ON THE MATHEMATICS TESTS

PHASE 3

ABSTRACT

The results of the tests have been reported in terms of proportions achieving competence. An analysis of such data enables one to identify areas in which students have done relatively well and those in which they have done poorly.

An estimate of sampling error for proportions has also been included as a cautionary measure for comparing individual school results against the Territory results. As a means of providing a more detailed information to schools, percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items in a test have also been reported.

4.1 TABLES

The following is a brief description of the tables in this chapter.

Table 1A shows percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items on each objective/test.

Table 1B shows percentages of students achieving competence (80% and more correct) on each objective/test and highlights the objectives in which more than 70% of students have shown competence and those in which less than 70% have shown competence.

Figure 1 shows overall performance in which areas of relative strength are shown against areas of relative weakness.

4.2 INTERPRETING STANDARDS

The level of competency or cutoff score recommended by the item-writing panel was 80% correct. The cutoff score is a measure of standard of performance. In addition to the cutoff score, the proportion of students achieving the cutoff score can be taken as a further measure of standards. However, what constitutes an acceptable proportion is a matter of judgement.

For the purpose of this report, the proportion of 70% of students achieving competence has been employed as a criterion for analysing performance on the objectives. This may be referred to as the 70-80 criterion.

In other words, the 70-80 criterion has been used as the benchmark, that is, 70% of students achieving competence fixed at 80% of items correct.

Those objectives in which more than 70% have shown competence are indicated in Table 1B.

4.3 COMPARING SCHOOL RESULTS WITH NORTHERN TERRITORY RESULTS

Table 1A may be used by schools to compare the results of their own students with the results obtained for the Territory.

In this table, estimates of sampling error are shown in the column for 80% correct. These estimates are based on the 95 per cent confidence limits.

For instance, for test M-A-1, in Table 1A, 71.4% of students achieved the competency level fixed at 80% correct. The sampling error for proportions estimated was plus or minus 7% points.

To determine whether differences are significant, any observed difference in this case should be at least 7% points. That is, in a school, the proportion achieving competence on test M-A-1 has to be more than 78% (71.4 + 7.0, rounded) for the result to be **significantly better** than the Territory performance.

Conversely, the proportion achieving competence has to be lower than 64% (71.4 - 7.0, rounded) for the result to be **significantly worse** than the Territory result.

For comparing school results with the Territory results, the graphs in the Appendix may be used. The 95% confidence limits are indicated in each graph.

Differences required for significance can be read directly from the graph.

4.4 OVERALL PERFORMANCE ON OBJECTIVES

From Table 1A, using the 80% correct as the competency level, the results may be summarised as follows:

Out of 28 objectives in Phase 3 Mathematics administered to Year 7 students:

On 16 objectives, more than 70% of students were competent.
On 7 objectives, between 50% and 70% of students were competent.
On 5 objectives, less than 50% of students were competent.

Overall, the results show that the majority of students in Year 7 were competent in most of the core objectives tested.

4.5 AREAS IN WHICH MORE THAN 70% HAVE ACHIEVED COMPETENCE

Using the 80% cutoff score as the competency level, Table 1B has been presented separately to highlight the objectives in which more than 70% have shown competence, and those in which less than 70% have shown competence.

The following lists the areas where more than 70% of students have achieved competence. The areas indicated below represent either single objectives or grouped objectives.

S-C-1	Identify position on a grid
M-A-1	Select relationship between mm, cm, m and km
M-B-1	Find area of 2-D figures by counting squares
M-B-2	Compare areas of 2-D figures by counting squares
M-C-2	Know 1 L = 1000 mL
M-O-2A	Identify symbols for length

N-B-3A	Add 4-digit numbers : equivalent digits
N-B-3B	Add 4-digit numbers : variable digits
N-B-7A	Solve word problems using addition
N-D-2	Multiply 2 digits by 2 digits
N-D-3	Divide, up to 4 digits, by single digits, with regrouping and remainders
N-D-4	Divide 3 digits by multiples of 10 (to 90) with regrouping and remainders
N-D-5	Multiply 3 digits (with up to 2 decimal places) by single digits
N-D-6	Divide 4 digits (with up to 2 decimal places) by single digits, no remainders
N-D-7A	Solve word problems using multiplication
N-E-9E	Show order of fractions with like denominators

4.6 AREAS IN WHICH LESS THAN 70% HAVE ACHIEVED COMPETENCE

From Table 1B, the following are the main areas in which less than 70% of students have achieved competence.

S-B-3	Measure angles between 0^{o} and 180^{o}
S-B-4	Find angle of given size
M-A-5	Convert from one unit to another (length) including use of decimals
M-D-4	Convert g into kg and vice versa
M-E-5	Find interval between two given times
M-O-2B	Identify symbols for mass (weight)
M-O-2C	Identify symbols for volume
N-B-7B	Solve word problems using subtraction
N-D-7B	Solve word problems using division
N-E-2	Identify place values in 5 digit numerals (with up to 3 decimal places)
N-E-6A	Show order relations between whole numbers
N-E-6B	Show order relations between numbers with decimals
N-F-2	Read simple bar, line graphs

TABLE 1A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
SPACE							
(1)	S-B-3	Measure angles between 0° and 180°	4		39.9 ±8		
(2)	S-B-4	Find angle of given size	4		50.1 ±8		
(3)	S-C-1	Identify position on a grid	5	83.0	83.0 ±6	63.2	63.2
MEASUREMENT							
(4)	M-A-1	Select relationship between mm, cm, m and km	8	84.1	71.4 ±7	45.8	45.8
(5)	M-A-5	Convert from one unit to another (length) including use of decimals	16	40.8	31.5 ±7	19.4	9.2
(6)	M-B-1	Find area of 2-D figures by counting squares	5	84.4	84.4 ±6	71.5	71.5
(7)	M-B-2	Compare areas of 2-D figures by counting squares	5	80.1	80.1 ±6	60.0	60.0
(8)	M-C-2	Know 1L = 1 000 mL	5	81.3	81.3 ±6	67.4	67.4
(9)	M-D-4	Convert g into kg and vice versa	9	16.6	8.3 ±4	0.7	0.7
(10)	M-E-5	Find interval between two given times	9	46.3	31.7 ±7	14.9	14.9
(11)	M-O-2A	Identify symbols for length	8	80.1	76.0 ±6	62.5	62.5
(12)	M-O-2B	Identify symbols for mass	4		65.6 ±7		
(13)	M-O-2C	Identify symbols for volume	4		38.2 ±8		

TABLE 1A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
NUMBER ADDITION AND SUBTRACTION							
(14)	N-B-3A	Add 4-digit numbers: equivalent digits	9	89.3	73.7 ±7	50.7	50.7
(15)	N-B-3B	Add 4-digit numbers: variable digits	4		79.1 ±6		
(16)	N-B-7A	Solve word problems using addition	10	83.7	80.4 ±6	71.6	49.0
(17)	N-B-7B	Solve word problems using subtraction	10	72.8	64.2 ±8	53.1	35.9
MULTIPLICATION AND DIVISION							
(18)	N-D-2	Multiply 2 digits by 2 digits	10	84.6	80.6 ±6	73.9	59.9
(19)	N-D-3	Divide, up to 4 digits, by by single digits, with regrouping and remainders	14	83.4	76.3 ±6	66.4	42.4
(20)	N-D-4	Divide 3 digits by multiples of 10 (to 90) with regrouping and remainders	10	76.6	71.6 ±7	62.9	43.2
(21)	N-D-5	Multiply 3 digits (with up to 2 decimal places) by single digits	15	79.4	73.4 ±7	51.8	29.8

TABLE 1A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
(22)	N-D-6	Divide 4 digits (with up to 2 decimal places) by single digits, no remainders	15	84.2	80.9 ±6	61.9	43.2
(23)	N-D-7A	Solve word problems using multiplication	10	77.4	71.3 ±7	60.4	35.4
(24)	N-D-7B	Solve word problems using division	10	75.3	67.7 ±7	58.1	36.6
(25)	N-E-2	Identify place values in 5 digit numerals (with up to 3 decimal places)	35	63.5	60.2 ±8	47.7	13.9
NUMBER RELATIONS							
(26)	N-E-6A	Show order relations between whole numbers	18	80.7	63.5 ±8	30.7	14.1
(27)	N-E-6B	Show order relations between numbers with decimals	8	74.4	63.4 ±8	37.8	37.8
(28)	N-E-9A	Show order of fractions with like denominators	13	83.6	80.4 ±6	75.5	65.0

TABLE 1B: PERCENTAGES ACHIEVING COMPETENCE: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
(1)	S-B-3	Measure angles between 0° and 180°	4	3	411	39.9*
(2)	S-B-4	Find angle of given size	4	3	365	50.1*
(3)	S-C-1	Identify position on a grid	5	4	288	83.0
MEASUREMENT						
(4)	M-A-1	Select relationship between mm, cm, m and km	8	6	402	71.4
(5)	M-A-5	Convert from one unit to another (length) including use of decimals	16	13	520	31.5*
(6)	M-B-1	Find area of 2-D figures by counting squares	5	4	404	84.4
(7)	M-B-2	Compare areas of 2-D figures by counting squares	5	4	367	80.1
(8)	M-C-2	Know 1L = 1000 mL	5	4	288	81.3
(9)	M-D-4	Convert g into kg and vice versa	9	7	301	8.3*
(10)	M-E-5	Find interval between two given times	9	7	410	31.7*

TABLE 1B: PERCENTAGES ACHIEVING COMPETENCE: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
MEASUREMENT						
(11)	M-O-2A	Identify symbols for length	8	6	533	76.0
(12)	M-O-2B	Identify symbols for mass	4	3	584	65.6*
(13)	M-O-2C	Identify symbols for volume	4	3	547	38.2*
NUMBER ADDITION AND SUBTRACTION						
(14)	N-B-3A	Add 4-digit numbers: equivalent digits	9	7	308	73.7
(15)	N-B-3B	Add 4-digit numbers: variable digits	4	3	306	79.1
(16)	N-B-7A	Solve word problems using addition	10	8	429	80.4
(17)	N-B-7B	Solve word problems using subtraction	10	8	371	64.2*
MULTIPLICATION AND DIVISION						
(18)	N-D-2	Multiply 2 digits by 2 digits	10	8	479	80.6
(19)	N-D-3	Divide, up to 4 digits, by single digits, with regrouping and remainders	14	11	295	76.3
(20)	N-D-4	Divide 3 digits by multiples of 10 (to 90) with regrouping and remainders	10	8	482	71.6

TABLE 1B: PERCENTAGES ACHIEVING COMPETENCE: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
(21)	N-D-5	Multiply 3 digits (with up to 2 decimal places) by single digits	15	12	523	73.4
(22)	N-D-6	Divide 4 digits (with up to 2 decimal places) by single digits, no remainders	15	12	278	80.9
(23)	N-D-7A	Solve word problems using multiplication	10	8	397	71.3
(24)	N-D-7B	Solve word problems using division	10	8	334	67.7*
RELATIONS AND NUMBER SYSTEMS						
(25)	N-E-2	Identify place values in 5 digit numerals (with up to 3 decimal places)	35	28	447	60.2*
(26)	N-E-6A	Show order relations between whole numbers	18	14	362	63.5*
(27)	N-E-6B	Show order relations between numbers with decimals	8	6	254	63.4*
(28)	N-E-9A	Show order of fractions with like denominators	13	10	408	80.4

[Note: * indicates less than 70% achieving competence]

S - B - 3	S - B - 4	S - C - 1	M - A - 1	M - A - 5	M - B - 1
M - B - 2	M - C - 2	M - D - 4	M - E - 5	M - O - 2A	M - O - 2B
M - O - 2C	N - B - 3A	N - B - 3B	N - B - 7A	N - B - 7B	N - D - 2
N - D - 3	N - D - 4	N - D - 5	N - D - 6	N - D - 7A	N - D - 7B
		N - E - 2	N - E - 6A	N - E - 6B	N - E - 9A

FIG 1: OVERALL PERFORMANCE MATHEMATICS PHASE 3

More than 70% achieved competence

Less than 70% achieved competence

CHAPTER FIVE

COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE MATHEMATICS PHASE 3

ABSTRACT

In this chapter, performance in 1986 was compared with that in 1985 on the results of the tests common to both years. The results seem to indicate that although there was only a limited number of tests where a comparison could be made, there were as many objectives where performance had improved as those where performance had declined. Some concern was expressed on the areas dealing with conversion of units.

5.1 INTERPRETING RESULTS

A comparison between 1985 and 1986 performance was carried out on the results from the same tests administered in both years. As the number of tests common to both years was restricted, this factor should be borne in mind when one interprets the outcome from the analysis.

In 1985, some principals and teachers commented that the testing program had imposed a heavy burden on resources in terms of time and the cost of photocopying test materials. Although this may be refuted on the grounds that testing the core (which is one of the main aims of the assessment program) was part of routine instruction, it was decided, nevertheless, to reduce the number of tests quite significantly in 1986.

The tests for 1986 were selected on the basis of the following considerations:

(1) the knowledge and skills involved were fundamental to future learning;

(2) the tests had sound statistical characteristics from previous analysis; and

(3) a few of them had poor results from the previous year's administration.

Therefore, bearing in mind that any comparison could only be done from an analysis of a limited number of tests common to both the 1985 and 1986 administrations, caution needs to be exercised in the interpretation of the outcome. A general statement regarding improvement or decline in standards may not be readily defensible. Rather, the findings would provide an indication on the areas where there seemed to have been a significant improvement or a significant decline.

5.2 COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE

The results from the 1985 and 1986 administrations were compared using the tests common to both administrations.

Table 1C shows percentages of students achieving competence in 1985 and 1986 on each test. Changes in performance that are statistically significant are indicated with an (*) beside them.

From Table 1C, the observations may be summarised as follows:

	Significant Improvement	Significant Decline	No Change
Number of objectives	5	5	5

5.3 AREAS WITH SIGNIFICANT IMPROVEMENT NOTED

The objectives or content areas that showed significantly better results in 1986 included the following:

(1) Divide, up to 4 digits, by single digits, with regrouping and remainders, (N-D-3).

(2) Divide 3 digits by multiples of 10 (to 90) with regrouping and remainders, (N-D-4).

(3) Divide 4 digits (with up to 2 decimal places) by single digits, no remainders, (N-D-6).

(4) Identify position on a grid, (S-C-1).

(5) Know 1 L = 1 000 mL, (M-C-2).

The results on the division tests were very encouraging as they were the more difficult tests in the number area. The data in Table 1C show that in 1986 proportions competent in the division objectives were between 70% and 80%. It is interesting to note too that the first two objectives mentioned above were also the only ones that had recorded a significant decline in 1985.

5.4 AREAS WITH SIGNIFICANT DECLINE NOTED

There were five objectives in which a significant drop in performance was observed. These were:

(1) Convert g into kg and vice versa, (M-D-4).

(2) Multiply 3 digits (with up to 2 decimal places) by single digits, (N-D-5).

(3) Select correct relation between mm, cm, m and km, (M-A-1).

(4) Convert from one unit to another (length) including use of decimals, (M-A-5).

(5) Compare areas of 2-D figures by counting squares, (M-B-2).

It appears that performance was weak in areas involving operations with decimal numbers as shown in the tests on conversion of units, i.e. M-D-4 and M-A-5. There was also a decline in performance in these areas.

The results also seem to indicate that in the remaining five areas, the changes in performance were not significant.

TABLE 1C: CHANGES IN PERFORMANCE: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
(1)	S-B-3	Measure angles between 0° and 180°	4	39.5	39.9	+0.4
(2)	S-C-1	Identify position on a grid	5	72.5	83.0	+10.5*
(3)	M-A-1	Select correct relation between mm, cm, m and km	8	86.0	71.4	-14.6*
(4)	M-A-5	Convert from one unit to another (length) including use of decimals	16	38.8	31.5	-7.3*
(5)	M-B-1	Find area of 2-D figures by counting squares	5	78.5	84.4	+5.9
(6)	M-B-2	Compare areas of 2-D figures by counting squares	5	86.2	80.1	-6.1*
(7)	M-C-2	Know 1 L = 1 000 mL	5	67.8	81.3	+13.5*
(8)	M-D-4	Convert g into kg and vice versa	9	34.4	8.3	-26.1*
(9)	M-E-5	Find interval between two given times	9	27.4	31.7	+4.3
(10)	N-D-2	Multiply 2 digits by 2 digits	10	81.0	80.6	-0.4
(11)	N-D-3	Divide, up to 4 digits, by single digits, with regrouping and remainders	14	64.4	76.3	+11.9*

TABLE 1C: CHANGES IN PERFORMANCE: MATHEMATICS PHASE 3

NO	CODE	OBJECTIVES	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
(12)	N-D-4	Divide 3 digits by multiples of 10 (to 90) with regrouping and remainders	10	62.4	71.6	+9.2
(13)	N-D-5	Multiply 3 digits (with up to 2 decimal places) by single digits	15	83.0	73.4	-9.6*
(14)1	N-D-6	Divide 4 digits (with up to 2 decimal places) by single digits, no remainders	15	70.9	80.9	+10.0*
(15)	N-E-2	Identify place values in 5 digit numerals (with up to 3 decimal places)	35	57.9	60.2	+2.3

CHAPTER SIX

RESULTS ON SELECTED TESTS AND ITEMS
MATHEMATICS PHASE 3

ABSTRACT

In this chapter, some tests have been selected for more detailed analysis. Where it seems of particular interest, a few test items have been reproduced. Test means and proportions passing an item were the performance indicators used.

6.1

Twenty eight separate tests in Phase 3 were administered to Year 7 students in 1986. Space would not permit discussion of all the tests used.

However, in order to provide some clues to the nature of the difficulty that some students seemed to have, some items have been reproduced. Both kinds of items have been reproduced : those that had high proportions of students passing as well as those that had high proportions failing.

Test means and standard deviations for the last three years have been included as a matter of interest.

6.2 SPACE

(1) Test P3SC1

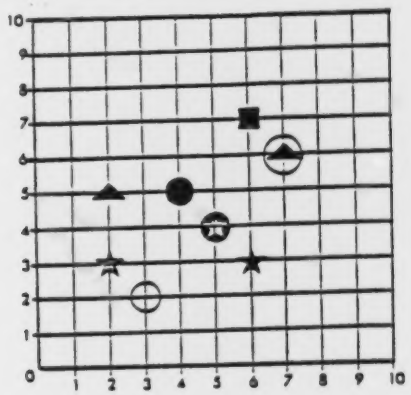
Objective: Given the coordinates of a point, identify its position, using the first quadrant only.

Results

	1984	1985	1986
MEANS	4.05	3.95	4.32
STANDARD DEVIATION	1.44	1.44	1.09

NUMBER OF ITEMS = 5

Examples



Item 1

Which symbol marks (3,2)?

A ☆ B ● C ✪ D ◯

% Correct = 82

Item 4

Which symbol marks (2,5)?

A ☆ B ▲ C ● D ★

% Correct = 84

Discussion

Average proportion passing an item was 86% compared with 79% in 1985. The results showed that this test was easy for most children. On every item, more than 80% got it right.

(2) Test P3SB3

Objective: Given an angle, between 0^{o} and 180^{o}, determine its size.

Results

	1984	1985	1986
MEANS	2.86	2.92	2.99
STANDARD DEVIATION	1.25	1.10	1.06

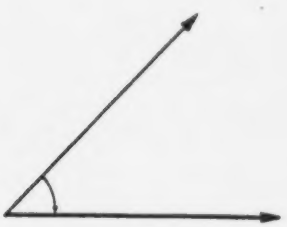
NUMBER OF ITEMS = 4

Examples

Items 1 and 4 have been selected for illustration.

Item 1

What is the size of this angle?



A 47^{o}
B 43^{o}
C 133^{o}
D 143^{o}

% Correct = 81

Item 4

What is the size of this angle?

A 35^{o}
B 38^{o}
C 142^{o}
D 145^{o}

% Correct = 63

Discussion

Average proportion passing an item was 75% compared with 73% in 1985. Measuring angles less than 90^{o} was easy for most children. However, more than 25% had trouble with two items, items 3 and 4 which required them to measure angles greater than 90% but less than 180^{o}.

6.3 MEASUREMENT

3(a) Test P3M02A

Objective: Given a unit of length in words, identify its appropriate symbol.

Example

Item 4

The distance from Alice Springs to Darwin is about 1 500 kilometres. Another correct way to write this is

A 1 500 k's
B 1 500 m
C 1 500 cm
D 1 500 km

% Correct = 86

3(b) Test P3M02B

Objective: Given a unit of weight in words, write its appropriate symbol.

Example

Item 3

A man weighs 70 kilograms. Another correct way to write this is:

A 70 K
B 70 Kg
C 70 kg
D 70 KG

% Correct = 70

3(c) Test P3M02C

Objective: Given a unit of volume in words, identify its appropriate symbol.

Examples

Item 1

This jug has 370 millilitres of milk in it.

Another correct way of writing this is

A 370 ML
B 370 ml
C 370 mL
D 370 Ml

% Correct = 41

Item 4

This water tank holds 5 000 litres of water.

Another correct way of writing this is

A	5 000 mL	C	5 000 ml
B	5 000 L	D	5 000 l

% Correct = 81

Discussion

Most students had no difficulty with tests M02A and M02B which involved units of length and weights respectively. In test M02C most had no trouble in identifying the correct symbol for litres, but a high proportion did not know the correct symbol for millilitres, as Item 1 in this test showed.

Similar results were observed in 1985. In both years more than 40% of children could not recognise the correct symbol for millilitres which is mL.

(4) Test P3MA5

Objective: Given length/distance in one unit, write the same measurement in another unit, including the use of decimals.

Results

	1984	1985	1986
MEANS	9.29	10.28	9.77
STANDARD OF DEVIATION	4.53	4.33	4.41
NUMBER OF ITEMS = 16			

Examples

Item 5

4 km = _____ m

A 40 000 m
B 400 000 m
C 400 m
D 4 000 m

% Correct = 77

Item 6

7 cm = _____ m

A 70 m
B 0.7 m
C 0.07 m
D 700 m

% Correct = 50

Item 9

0.6 cm = _____ mm

A 0.06 mm
B 6 mm
C 60 mm
D 600 mm

% Correct = 45

Discussion

This 16-item test turned out to be one of the most difficult. Average proportion passing an item was 61% compared with 64% the previous year. 40% of children failed to give the correct response to 10 out of the 16 items. The most difficult were Items 3 and 9 in which the student had to change a decimal unit.

(5) Test P3MC2

Objective: Know 1 L = 1 000 mL

Results

	1984	1985	1986
MEANS	3.80	3.88	4.23
STANDARD DEVIATION	1.61	1.63	1.38

NUMBER OF ITEMS = 5

Example

Item 1

1 L = _____ mL

A 0.001 mL
B 10 mL
C 100 mL
D 1 000 mL

% Correct = 88

Discussion

In this 5-item test on the equivalence of one litre and 1 000 millilitres, average proportion passing an item was 85% compared with 78% in 1985. It may be concluded that the equivalence 1 L = 1 000 mL or 1 000 mL = 1 L was correctly recalled by most children.

(6) Test P3MD4

Objective: Given mass in one unit, write the same mass in another unit or units, including the use of decimals.

Results

	1984	1985	1986
MEANS	5.67	6.15	4.99
STANDARD DEVIATION	2.56	2.12	1.83

NUMBER OF ITEMS = 9

Examples

Items 1 and 3 were found to be easy by most students, but not Items 6, 7, 8 or 9.

Item 1

1 050 g = _____ kg

A 1 kg 50 g
B 10 kg 5 g
C 10 kg 50 g
D 105 kg 0 g

% Correct = 81

Item 3

2 kg 550 g = _____ g

A 552 g
B 750 g
C 2 550 g
D 20 550 g

% Correct = 84

Item 6

4.83 kg is the same as

A 4 083 g
C 483 kg
B 4 kg 830 g
D 4 kg 83 g

% Correct = 30

Item 7

A bag of sugar weighs 2 kg. The mass is the same as

A	0.2000 g	B	2 g
C	200 g	D	2 000 g

% Correct = 24

Item 8

8 kg 75 g is the same as

A	.875 g	B	8 075 g
C	8 705 g	D	8 750 g

% Correct = 22

Item 9

7 kg 362 g is the same as

A	7.362 g	B	73.62 g
C	736.2 g	D	7 362 g

% Correct = 22

Discussion

Average proportion passing an item was 55% compared with 63% in 1985. It appears that a high proportion had difficulty with conversion from one unit to another and many had trouble in dealing with decimal conversion.

This particular test has consistently produced poor performance in the last three years as indicated by the figures for the means.

(7) P3ME5

Objective: Calculate the time interval between two given times, where one source is a timetable and the other a stated time.

Results

	1984	1985	1986
MEANS	5.74	6.08	6.06
STANDARD DEVIATION	2.28	1.99	2.26

NUMBER OF ITEMS = 9

Examples

'TODAY'S AFTERNOON PROGRAM OF TV SPORT'

TV Program	Time
Baseball Game	2:00 pm
Tennis Game	3:00 pm
Football Game	4:00 pm
Hockey Game	5:00 pm
Finish	6:00 pm

Item 3

It is 8:00 pm. How long has it been since the baseball game started?

A 10 hours
B 6 hours
C 7 hours
D 4 hours

% Correct = 83

Item 5

It is 9:15 am. How long will it be until the baseball game finishes?

A 6 h 45 min
B 12 h 45 min
C 5 h 45 min
D 4 h 45 min

% Correct = 51

Discussion

As the above two examples illustrate, many students had trouble in computing time intervals in which both hours and minutes were involved. Average proportion passing on item was 67% compared with 64% in 1985.

6.4 NUMBER

(8) Test P3NB7B

Objective: Solve word problems using subtraction skills.

Results

	1984	1985	1986
MEANS	-	8.29	7.59
STANDARD DEVIATION	-	2.43	2.83

NUMBER OF ITEMS = 10

Examples

Item 5

What change would you get from \$10.00, if you paid \$8.35 for a pair of shoes?

A \$2.65 B \$1.65 C \$1.75 D \$2.75

% Correct = 77

Item 6

The price of a bike in a sale was marked down from \$174 to \$129. How much would you save if you bought the bike?

A \$46 B \$56 C \$55 D \$45

% Correct = 79

Discussion

Average proportion passing an item was 76%. Most children could recognise the problems as a subtraction task and compute the answers correctly.

(9) Test P3ND3

Objective: Divide whole numbers to 4-digits by single digit numbers with regrouping and whole remainders.

Results

	1984	1985	1986
MEANS	12.16	10.91	11.94
STANDARD DEVIATION	2.73	3.95	3.11

NUMBER OF ITEMS = 14

Examples

Item 5

$3\overline{)798}$

A 232 r2 B 299 C 260 r6 D 266

% Correct = 85

Item 7

$5\overline{)538}$

A 101 r3 B 106 r8 C 107 r3 D 17 r3

% Correct = 61

Discussion

Average proportion passing an item was 85% compared with 78% in 1985. The results seem to suggest that most children could do simple division with single digit divisors. Except for Item 7, most found the task quite easy.

(10) Test P3ND5

Objective: Multiply 3-digit numbers with up to 2 decimal places, including money, by single digit numbers.

Results

	1984	1985	1986
MEANS	11.64	13.21	12.22
STANDARD DEVIATION	3.75	2.41	3.64

NUMBER OF ITEMS = 15

Examples

Item 5

20.6 mm
x 2

A 41.2 mm B 40.12 mm C 4.12 mm D 42.8 mm

% Correct = 66

Item 15

59.6
x 3

A 157.8 B 178.8 C 15.78 D 17.88

% Correct = 60

Item 9

$1.56

x 3

A $3.58 B $4.68 C $4.58 D $4.89

% Correct = 85

Discussion

Average proportion passing an item was 81% compared with 88% in 1985. Although there was a drop in performance in 1986, the results still showed that 80% of children could multiply 3 digit numbers (including decimals) by single digits with accuracy.

(11) Test P3ND6

Objective: Divide 4 digit numbers with up to 2 decimal places, including money, by single digit numbers, with no remainders.

Results

	1984	1985	1986
MEANS	11.89	11.41	12.93
STANDARD DEVIATION	3.81	5.10	3.01

NUMBER OF ITEMS = 15

Examples

Item 6

$3\overline{)\$57.09}$

A $29.30 B $29.03 C $19.03 D $19.30

% Correct = 85

Item 12

$4\overline{)905.6}$

A 22.64 B 226.4 C 26.4 D 2.264

% Correct = 87

Discussion

Results on the test improved in 1986 with an average of 86% passing an item compared with 76% in 1985. This was another test on number that most children found easy.

(12) Test P3NE2

Objective: Identify place value and meaning of each digit in 5-digit numerals with up to 3 decimal places.

Results

	1984	1985	1986
MEANS	19.04	24.48	25.59
STANDARD DEVIATION	12.37	11.12	10.34

NUMBER OF ITEMS = 35

Examples

Item 5

What digit is in the units/ones place in the number 6 420.9?

A 0 B 2 C 9 D 6

% Correct = 83

Item 8

What digit is in the tens place in the number 8 462.7?

A 4 B 2 C 7 D 6

% Correct = 75

Item 14

What digit is in the hundredths place in the number 361.47.

A 4 B 7 C 1 D 3

% Correct = 71

Discussion

Average proportion passing an item was 73% compared with 70% in 1985. Less than 30% had difficulty in dealing with place values. A criticism that could be levelled at this test was that there were too many items. 10 items should be sufficient to measure this particular objective.

CHAPTER SEVEN

ANALYSIS OF PERFORMANCE ON THE MATHEMATICS TESTS

PHASE 2

ABSTRACT

The results of the tests have been reported in terms of proportions achieving competence. An analysis of such data enables one to determine areas where students have done relatively well and those in which they have done poorly.

An estimate of sampling error for proportions has also been included as a cautionary measure for comparing individual school results against the Territory results. As a means of providing more detailed information to schools, percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items in a test have also been reported.

7.1 TABLES

The following is a brief description of the tables in this chapter.

Table 2A shows percentages of students responding correctly to more than 70%, 80%, 90% and 100% of items on each objective/test.

Table 2B shows percentages of students achieving competence (80% and more correct) on each objective/test and highlights the objectives in which more than 70% of students have shown competence and those in which less than 70% have shown competence.

Figure 2 shows overall performance in which areas of relative strength are shown against areas of relative weakness.

7.2 INTERPRETING STANDARDS

The level of competency or cutoff score recommended by the item-writing panel was 80% correct. The cutoff score is a measure of standard of performance. In addition to the cutoff score, the proportion of students achieving the cutoff score can be taken as a further measure of standards. However, what constitutes an acceptable proportion is a matter of judgement.

For the purpose of this report, the proportion of 70% of students achieving competence has been employed as a criterion for analysing performance on the objectives. This may be referred to as the 70-80 criterion.

In other words, the 70-80 criterion has been used as the benchmark, that is, 70% of students achieving competence fixed at 80% of items correct.

Those objectives in which more than 70% have shown competence as indicated in Table 2B.

7.3 COMPARING SCHOOL RESULTS WITH NORTHERN TERRITORY RESULTS

Table 2A may be used by schools to compare the results of their own students with the results obtained for the Territory.

In this table, estimates of sampling error are shown in the column for 80% correct. These estimates are based on the 95 per cent confidence limits.

For instance, for test M-O-1, in Table 2A, 73.3% of students achieved the competency level fixed at 80%. The sampling error for proportions estimated was plus or minus 7% points.

To determine whether differences are significant, any observed difference in this case should be at least 7% points. That is, in a school, the proportion achieving competence on test M-O-1 has to be more than 80% (73.3 + 7.0, rounded) for the result to be **significantly better** than the Territory performance.

Conversely, the proportion achieving competence has to be lower than 66% (73.3 - 7.0, rounded) for the result to be **significantly worse** than the Territory result.

For comparing school results with the Territory results, the graphs in the Appendix may be used. The 95% confidence limits are indicated in each graph.

Differences required for significance may be read directly from the graph.

7.4 OVERALL PERFORMANCE ON OBJECTIVES

From Table 2B, using the 80% correct as the competency level, the results may be summarised as follows:

Out of 33 objectives in Phase 2 Mathematics administered to Year 5 students:

On 20 objectives, more than 70% of students were competent.
On 9 objectives, between 50% and 70% of students were competent.
On 4 objectives, less than 50% of students were competent.

Overall, the results show that the majority of students in Year 5 were competent in most of the core objectives tested.

7.5 AREAS IN WHICH MORE THAN 70% HAVE ACHIEVED COMPETENCE

Using the 80% cutoff score as the competency level, Table 2B has been presented separately to highlight the objectives in which more than 70% have shown competence, and those in which less than 70% have shown competence.

The following lists the areas where more than 70% of students have achieved competence. The areas indicated below represent either single objectives or grouped objectives.

M-O-1	Recall simple equivalence
M-B-1	Compare areas
M-D-5	Read a calendar
M-D-7B	Read time (digital)

N-A-2A	Add two numbers no regrouping
N-A-3B	Subtract multiples of 10 from number less than 100.
N-C-1A	Recall multiplication (tables)
N-D-1	Multiply single digit number by multiples of 10 (to 90)
N-D-2	Multiply two digit numbers by single digits
N-D-4	Divide money, by single digits, without remainders
N-E-1	Write largest/smallest number
N-E-2	Identify place values
N-E-4	Express given number in words
N-E-5A	Show order relations using symbols
N-E-5B	Arrange a set of numbers in correct order
N-E-7A	Recognise odd, even and square numbers
N-F-2	Read bar graph

7.6 AREAS IN WHICH LESS THAN 70% HAVE ACHIEVED COMPETENCE

From Table 2B, the following are the main areas in which less than 70% of students have achieved competence.

S-A-2	Use ordered pairs
M-O-2	Identify symbols for units
M-D-7A	Read time (colloquial)
N-A-2B	Add two numbers, with regrouping
N-A-3A	Subtract single digit number from number less than 100
N-B-2	Subtract numbers up to 3 digits
N-B-3	Add numbers up to 3 digits with decimals
N-B-4	Subtract 3 digit numbers with decimals
N-C-1B	Recall division (tables)
N-D-3	Divide two digit numbers by single digits, with and without remainders.
N-E-6A	Round to nearest 10, 100.

TABLE 2A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 2

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
SPACE							
(1)	S-A-2	Use ordered pairs	8	56.0	47.1 +-8	32.3	32.3
MEASUREMENT							
(2)	M-0-1	Recall simple equivalence	14	81.5	73.3 +-7	60.3	40.1
(3)	M-0-2	Identify symbols	10	79.1	68.8 +-7	55.1	28.8
(4)	M-B-1A	Compare areas to find largest area	5	90.4	90.4 +-5	75.2	75.2
(5)	M-B-1B	Compare areas to find smallest area	5	89.7	89.7 +-5	79.9	79.9
(6)	M-B-1C	Compare areas to find two shapes with same area	5	82.9	82.8 +-6	73.3	73.3
(7)	M-D-5	Read a calendar	6	77.4	77.4 +-6	61.5	61.5
(8)	M-D-7A	Read time (colloquial)	8	56.8	39.7 +-8	18.4	18.4
(9)	M-D-7B	Read time (digital)	5	74.7	74.7 +-7	55.9	55.9
NUMBER							
(10)	N-A-2A	Add two numbers, no regrouping	12	95.3	90.9 +-5	83.3	62.9
(11)	N-A-2B	Add two numbers with regrouping	29	73.3	62.0 +-8	50.2	19.6
(12)	N-A-3A	Subtraction of single digit number from number less than 100	6	69.5	69.5 +-7	46.3	46.3
(13)	N-A-3B	Subtraction multiples of 10 from number less than 100	6	87.5	87.5 +-5	69.7	69.7
(14)	N-B-2	Subtract numbers up to 3 digits	9	69.6	57.9 +-8	38.0	38.0

TABLE 2A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 2

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
NUMBER (Cont'd)							
(15)	N-B-3	Add numbers up to 3 digits with decimals	27	64.2	51.8 +/-8	29.1	6.3
(16)	N-B-4	Subtract 3 digit numbers with decimals	12	57.0	51.9 +/-8	43.0	23.8
(17)	N-C-1A	Recall basic facts: multiplication (tables)	19	85.7	74.6 +/-7	49.7	22.3
(18)	N-C-1B	Recall basic facts: division (tables)	10	69.3	61.3 +/-8	50.0	32.1
(19)	N-D-1	Multiply single digit number by multiples of 10 (to 90)	10	77.8	74.5 +/-7	66.0	49.5
(20)	N-D-2	Multiply two digit numbers by single digits	15	75.8	71.7 +/-7	53.4	30.3
(21)	N-D-3	Divide two digit numbers by single digits, with and without remainders	26	52.9	44.4 +/-8	27.1	7.5
(22)	N-D-4	Divide money by single digits, without remainders	5	77.4	77.4 +/-6	56.2	56.2
RELATIONS							
(23)	N-E-1	Write largest/smallest number given 3 digits	5	85.4	85.4 +/-5	57.8	57.8
(24)	N-E-2A	Identify place values	5	83.0	83.0 +/-6	77.1	77.1
(25)	N-E-2B	Identify face values	5	88.1	88.1 +/-5	77.7	77.7
(26)	N-E-2C	Identify total values	5	62.2	62.2 +/-8	35.0	35.0
(27)	N-E-4	Express given number in words	8	94.6	92.1 +/-5	73.4	75.4
(28)	N-E-5A	Show order relations using symbols	15	77.1	72.0 +/-7	56.7	35.2

TABLE 2A: PERCENTAGES AT DIFFERENT CUTOFFS: MATHEMATICS PHASE 2

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING 70%	80%	90%	100%
NUMBER (Cont'd)							
(29)	N-E-5B	Arrange a set of numbers in correct order	9	80.5	77.0 ±6	55.1	55.1
(30)	N-E-6A	Round to nearest 10	8	73.0	56.0 ±8	36.5	36.5
(31)	N-E-6B	Round to nearest 100	7	52.1	40.8 ±8	22.9	22.9
(32)	N-E-7A	Recognise odd, even and square numbers to 100	6	81.3	81.3 ±6	66.5	66.5
(33)	N-F-2	Read bar graph	5	88.8	88.8 ±5	69.7	69.7

TABLE 2B: PERCENTAGES ACHIEVING COMPETENCE: MATHEMATICS PHASE 2

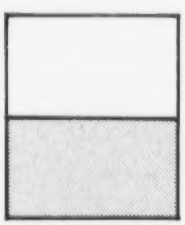
NO	CODE	OBJECTIVES	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
SPACE						
(1)	S-A-2	Use ordered pairs	8	6	520	47.1*
MEASUREMENT						
(2)	M-0-1	Recall simple equivalence	14	11	292	73.3
(3)	M-0-2	Identify symbols for units	10	8	503	68.8*
(4)	M-B-1A	Compare areas to find largest area	5	4	375	90.4
(5)	M-B-1B	Compare areas to find smallest area	5	4	378	89.7
(6)	M-B-1C	Compare areas to find two shapes with same area	5	4	378	82.8
(7)	M-D-5	Read a calendar	6	5	239	77.4
(8)	M-D-7A	Read time (colloquial)	8	6	380	39.7*
(9)	M-D-7B	Read time (digital)	5	4	379	74.7
NUMBER						
(10)	N-A-2A	Add two numbers, no regrouping	12	10	275	90.9
(11)	N-A-2B	Add two numbers, with regrouping	29	23	255	62.0*
(12)	N-A-3A	Subtract single digit number from number less than 100	6	5	354	69.5
(13)	N-A-3B	Subtract multiples of 10 from number less than 100	6	5	409	87.5
(14)	N-B-2	Subtract numbers up to 3 digits	9	7	171	57.9*
(15)	N-B-3	Add numbers up to 3 digits with decimals	27	22	461	51.8*
(16)	N-B-4	Subtract 3 digit numbers with decimals	12	10	428	51.9*

TABLE 2B: PERCENTAGES ACHIEVING COMPETENCE: MATHEMATICS PHASE 2

NO	CODE	OBJECTIVES	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
NUMBER (Cont'd)						
(17)	N-C-1A	Recall multiplication (tables)	19	15	503	74.6
(18)	N-C-1B	Recall division (tables)	10	8	504	61.3*
(19)	N-O-1	Multiply single digit number by multiples of 10 (to 90)	10	8	388	74.5
(20)	N-O-2	Multiply two digit numbers by single digits	15	12	442	71.7
(21)	N-O-3	Divide two digit numbers by single digits, with and without remainders	26	21	520	44.4*
(22)	N-O-4	Divide money by single digits, without remainders	5	4	477	77.4
RELATIONS						
(23)	N-E-1	Write largest/smallest number given 3 digits	5	4	486	85.4
(24)	N-E-2A	Identify place values	5	4	471	83.0
(25)	N-E-2B	Identify face values	5	4	462	88.1
(26)	N-E-2C	Identify total values	5	4	466	62.2*
(27)	N-E-4	Express given number in words	8	6	390	92.1
(28)	N-E-5A	Show order relations using symbols	15	12	471	72.0
(29)	N-E-5B	Arrange a set of numbers in correct order	9	7	483	77.0
(30)	N-E-6A	Round to nearest 10	8	6	482	56.0*
(31)	N-E-6B	Round to nearest 100	7	6	480	40.8*
(32)	N-E-7A	Recognise odd, even and square numbers to 100	6	5	251	81.3
(33)	N-F-2	Read bar graph	5	4	241	88.8

S - A - 2	M - O - 1	M - O - 2	M - B - 1A	M - B - 1B	M - B - 1C
M - D - 5	M - D - 7A	M - D - 7B	N A - 2A	N - A - 2B	N - A - 3A
N - A - 3B	N - B - 2	N - B - 3	N - B - 4	N - C - 1A	N - C - 1B
N - D - 1	N - D - 2	N - D - 3	N - D - 4	N - E - 1	N - E - 2A
N - E - 2B	N - E - 2C	N - E - 4	N - E - 5A	N - E - 5B	N - E - 6A
			N - E - 6B	N - E - 7A	N - F - 2

FIG 2: OVERALL PERFORMANCE MATHEMATICS PHASE 2



More than 70% achieved competence

Less than 70% achieved competence

CHAPTER EIGHT

COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE
MATHEMATICS PHASE 2

ABSTRACT

There were twenty four tests common to both 1985 and 1986 administrations. Using the results on these tests, performance in 1986 was compared with that in 1985. Declines seem to have outnumbered gains, but in one-third of the areas tested, changes in performance between 1985 and 1986 were only marginal and therefore not significant.

8.1 SUMMARY OF CHANGES IN PERFORMANCE BETWEEN 1985 AND 1986

Table 2C shows percentages of students achieving the competency score of 80% correct on the twenty four individual tests in Phase 2 administered in 1985 and 1986. Changes in performance are shown in the column on the extreme right. A plus (+) sign indicates a gain, a minus sign (-) indicates a decline. Changes in performance which were statistically significant are shown with an asterisk (*).

The changes in performance shown in Table 2C may be summarised as follows:

	SIGNIFICANT IMPROVEMENT	SIGNIFICANT DECLINE	NO CHANGE
NUMBER OF OBJECTIVES	2	13	9

The results on nine objectives did not show any significant change. However, the results on thirteen objectives showed a significant decline compared with the results obtained the previous year. Performance on two objectives improved over that of the previous year.

8.2 AREAS WITH SIGNIFICANT IMPROVEMENT NOTED

Only two objectives had significantly better results in 1986 and these were:

(1) Compare areas to find largest area, (M-B-1A).
(2) Compare areas to find smallest area, (M-B-1B).

Performance on the above objectives was already at a reasonably high level in 1985.

8.3 AREAS WITH SIGNIFICANT DECLINE NOTED

There were thirteen objectives in which a significant drop in performance was observed. These were:

(1) Use ordered pairs, (S-A-2).
(2) Identify symbols for units, (M-O-2).
(3) Read a calendar, (M-D-5).
(4) Read time (colloquial), (M-D-7A).
(5) Add numbers up to 3 digits with decimals, (N-B-3).
(6) Multiply single digit number by multiples of 10 (to 90), (N-D-1).
(7) Multiply two digit numbers by single digits, (N-D-2).
(8) Divide two digit numbers by single digits, with and without remainders, (N-D-3).
(9) Divide money by single digits, without remainders, (N-D-4).
(10) Identify place values, (N-E-2A).
(11) Write total values, (N-E-2C).
(12) Show order relations using symbols, (N-E-5A).
(13) Arrange a set of numbers in correct order, (N-E-5B).

Results on the remaining nine objectives showed only marginal gains or losses, but statistically, these were not significant.

TABLE 2C: CHANGES IN PERFORMANCE: MATHEMATICS PHASE 2

NO	CODE	TESTS	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
SPACE						
1	S-A-2	Use ordered pairs	8	64.8	47.1	-17.7*
MEASUREMENT						
2	M-0-2	Identify symbols for units	10	80.0	68.8	-11.2*
3	M-B-1A	Compare areas to find largest area	5	85.3	90.4	+5.1*
4	M-B-1B	Compare areas to find smallest area	5	82.3	89.7	+7.4*
5	M-B-1C	Compare areas to find two shapes with same area	5	79.9	82.8	+2.9
6	M-D-5	Read a calendar	6	87.0	77.4	-9.6*
7	M-D-7A	Read time (colloquial)	8	50.8	39.7	-11.1*
8	M-D-7B	Read time (digital)	5	76.7	74.7	-2.0
NUMBER						
9	N-B-2	Subtract numbers up to 3 digits	9	55.7	57.9	+2.2
10	N-B-3	Add numbers up to 3 digits with decimals	27	75.1	51.8	-23.3*
11	N-B-4	Subtract 3 digit numbers with decimals	12	58.3	51.9	-6.4
12	N-D-1	Multiply single digit number by multiples of 10 (to 90)	10	87.4	74.5	-12.9*
13	N-D-2	Multiply two digit numbers by single digits	15	83.0	71.7	-11.3*
14	N-D-3	Divide two digit numbers by single digits, with and without remainders	26	59.6	44.4	-15.2*
15	N-D-4	Divide money by single digits, without remainders	5	87.1	77.4	-9.7*

TABLE 2C: CHANGES IN PERFORMANCE: MATHEMATICS PHASE 2

NO	CODE	TESTS	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
RELATIONS						
16	N-E-1	Write largest/smallest number given 3 digits	5	88.6	85.4	-3.2
17	N-E-2A	Identify place values	5	89.4	83.0	-6.4*
18	N-E-2B	Identify face values	5	89.9	88.1	-1.8
19	N-E-2C	Write total values	5	71.9	62.2	-9.7*
20	N-E-5A	Show order relations using symbols	15	84.3	72.0	-12.3*
21	N-E-5B	Arrange a set of numbers in correct order	9	87.1	77.0	-10.1*
22	N-E-6A	Round to nearest 10	8	54.4	56.0	+1.6
23	N-E-6B	Round to nearest 100	7	41.4	40.8	-0.6
24	N-F-2	Read bar graph	5	88.1	88.8	+0.7

CHAPTER NINE

RESULTS ON SELECTED TESTS AND ITEMS
MATHEMATICS PHASE 2

ABSTRACT

Some of the tests from Phase 2 administered to Year 5 students in 1986 have been selected for a closer examination. Within each test, a few items have been reproduced to give the reader some idea of the item content. The proportion of students passing an item is a very useful indicator of the 'difficulty' of the item, and this index is reported on each item reproduced. As a matter of interest, the test means for 1984, 1985 and 1986 have also been shown.

9.1

In 1986, thirty-three separate tests in Phase 2 were administered to Year 5 students. Space would not permit discussion of all the items used.

However, in order to provide the reader with some idea of the test content, some items have been reproduced. Both hard and easy items have been chosen for the purpose of the discussion.

Test means and standard deviations for the last three years have been included as a matter of interest, wherever possible.

9.2 SPACE

(1) Test P2SA2

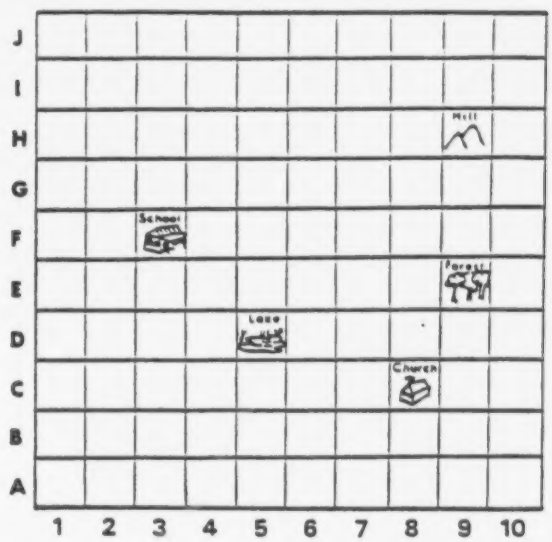
Objective: Use ordered pairs to describe positions of squares on a grid.

Results

	1984	1985	1986
MEANS	6.52	6.57	5.78
STANDARD DEVIATION	1.89	1.98	2.13

NUMBER OF ITEMS = 8

Examples



Item 1

Which ordered pair shows the position of the school?

A (6,C) B (4,F) C (3,F) D (8,C)

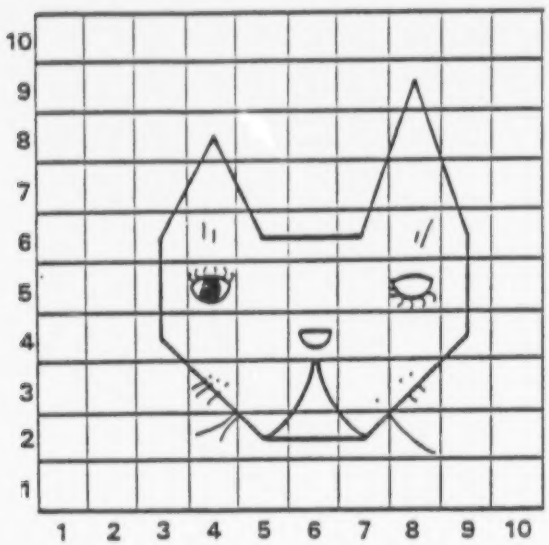
% Correct = 88

Item 4

Which ordered pair shows the position of the lake?

A (4,F) B (4,E) C (5,D) D (6,D)

% Correct = 88



Item 6

Which ordered pair shows the position of the nose?

A (4,4) B (4,6) C (6,4) D (6,6)

% Correct = 57

Item 7

One eye is open, which one?

A (5,4) B (4,5) C (4,4) D (5,5)

% Correct = 54

Discussion

Average proportion passing an item on this test was 72% compared with 82% in 1985. Most students did well on the first four items but rather poorly on the last four items (please refer to the original test).

In both years, the pattern was similar. It might be that the reason for this was the difficulty some students had with the second drawing and the language used.

9.3 MEASUREMENT

(2) Test P2M02

Objective: Identify the following symbols:

m, cm, L, g and kg.

Item 1

A jetty is 34 m long. The symbol underlined means

A millimetres B millilitres C metres D milligrams

% Correct = 90

Item 6

The car is 3 metres long. Another correct way to write 3 metres is

A 3 mtrs B 3 m's C 3M D 3 m

% Correct = 63

Item 2

The blackboard is 641 cm. The symbol underlined means

A centigrade B cents C celcius D centimetres

% Correct = 94

Item 7

The exercise book is 36 centimetres long. Another correct way to write 36 centimetres is

A 36 cm B 36 Cm C 36 CM D 36 ctm

% Correct = 79%

Item 8

The milk jug contains 2 litres of milk. Another correct way to write 2 litres is

A 2 l B 2 L C 2 lts D 2 l's

% Correct = 56

Discussion

Average proportion passing an item was 80% compared with 86% in 1985. The results showed that most Year 5 children were able to identify the symbols for the units used. Two interesting observations were noted.

(i) Items where the symbol was given and the unit was required appeared to be easier than items in which the unit was given and the symbol was asked. Items 1, 2, 6 and 7 illustrate this.

(ii) More than 40% of children did not know the correct symbol for litres, as item 8 shows.

(3) Test P2MD7A

Objective: Given four clock faces, read the time (colloquial).

Results

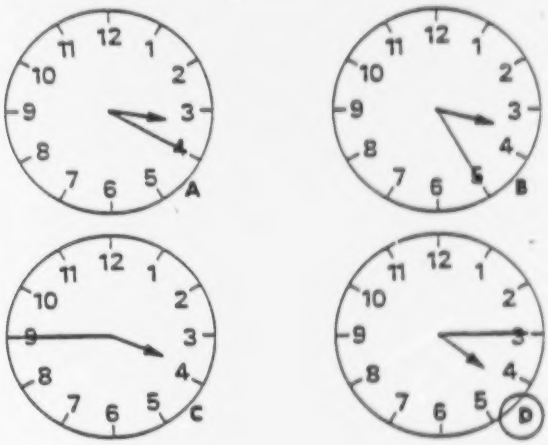
	1984	1985	1986
MEANS	-	6.13	5.64
STANDARD DEVIATION	-	1.63	1.87

NUMBER OF ITEMS = 8

Examples

Item 3

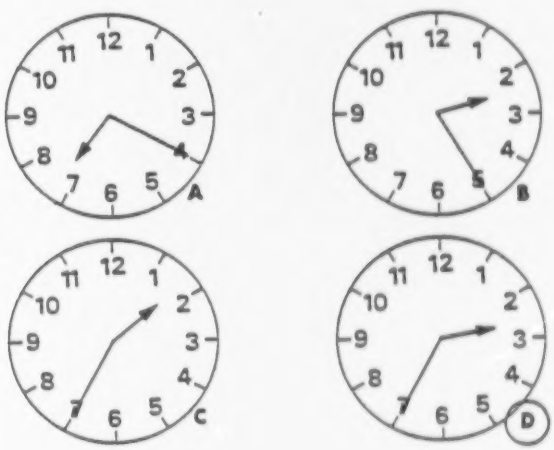
Which one of these clocks shows a quarter past four?



% Correct = 78%

Item 4

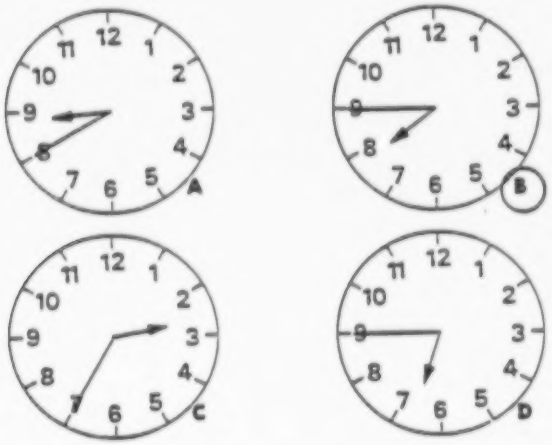
Which one of these clocks shows twenty five to three?



% Correct = 58

Item 7

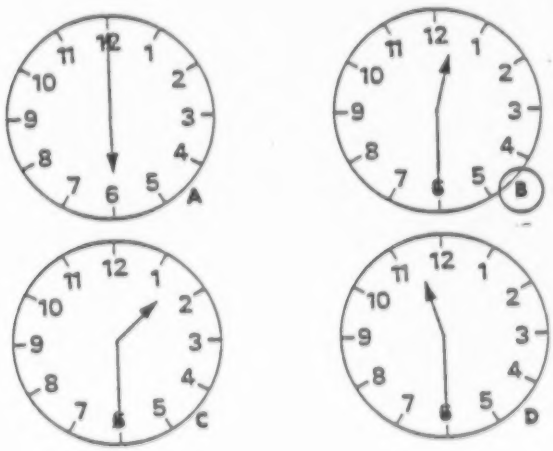
Which one of these clocks shows 7:45 h?



% Correct = 53

Item 6

Which one of these clocks shows 12:30 h?



% Correct = 49

Discussion

The test required students to read the time on a round clockface. Average proportion successful on an item was 71% compared with 77% in 1985. As the examples show, quite a high proportion of children had trouble with items 4, 6 and 7. Most did not have any difficulty in matching the correct time with the times stated colloquially if the minute hand was on the hour, half hour or quarter hour.

9.4 NUMBER

(4) Test P2NB3

Objective: Add up to three-digit numbers with one or two decimal places.

Results

	1984	1985	1986
MEANS	22.54	22.87	19.05
STANDARD DEVIATION	5.46	4.88	7.33

NUMBER OF ITEMS = 27

Examples

Item 5

63.7
+2.8
———

A 65.4 B 66.4 C 66.5 D 6.5

% Correct = 80

Item 10

$4.72
$1.19
———

A $5.91 B $5.81 C $591.00 D $6.91

% Correct = 75

Item 22

$2.54
$3.16
$1.27
———

A $69.70 B $6.87 C $6.97 D $7.97

% Correct = 61

Item 13

$1.56
$2.79
———

A $3.25 B $3.35 C $4.25 D $4.35

% Correct = 58

Item 27

$4.67
$0.99
$3.78
———

A $8.44 B $9.24 C $7.44 D $9.44

% Correct = 50

Discussion

The test required students to add two rows or three rows of numbers with decimals. Average proportion passing an item was 71% compared with 85% in 1985.

From the results, it seems that quite a high proportion of students found the task difficult particularly where carrying from two columns was involved. Items 13 and 27 illustrate this.

(5) Test P2NB4

Objective: Subtract 3 digit numbers with one or two decimal places, regrouping in the first column only.

Results

	1984	1985	1986
MEANS	9.59	9.40	8.46
STANDARD DEVIATION	2.69	2.64	3.39

NUMBER OF ITEMS = 12

Examples

Item 2

93.4
-42.3
____

A 51.1 B 511.1 C 51.7 D 51.11

% Correct = 84

Item 5

94.0
-32.6
____

A 61.4 B 62.4 C 61.6 D 62.6

% Correct = 64

Item 6

75.2
-30.6
____

A 45.6 B 45.4 C 446 D 44.6

% Correct = 57

Item 11

6.90
-0.71
____

A 6.19 B 6.29 C 6.21 D 0.21

% Correct = 58

Discussion

Average proportion passing an item was 70% compared with 78% in 1985. Items 1, 2, 3, 7, 8 9 which did not require regrouping showed a much higher success rate compared with those which required regrouping as in items 5, 6, 10, 11 and 12. It appears that quite a significant proportion of students had some difficulty with subtraction tasks which involved regrouping. The same difficulty was observed in 1985.

(6) Test P2ND2

Objective: Multiply two-digit numbers by single digits with or without regrouping.

Results

	1984	1985	1986
MEANS	13.18	13.11	11.95
STANDARD DEVIATION	3.32	3.19	3.96

NUMBER OF ITEMS = 15

Examples

Item 6

15 kg
x 6
——

A 21 kg B 85 kg C 90 kg D 156 kg

% Correct = 86

Item 7

17 cm
x 6
___

A 102 cm B 23 cm C 176 cm D 112 cm

% Correct = 79

Item 9

$75
x 9
____

A $84 B $684 C $694 D $675

% Correct = 78

Item 10

75
x 8
___

A 5 640 B 758 C 600 D 560

% Correct = 60

Item 15

80
x 5
___

A 405 B 400 C 85 D 135

% Correct = 71

Discussion

Average proportion passing an item was 80% compared with 87% in 1985. Despite the drop in performance on this test, the success rate was still high. It may be concluded that most students were able to multiply two-digit numbers by single digits with or without regrouping.

(7) Test P2ND3

Objective: Divide two digit numbers by single digits, with and without remainders.

Results

	1984	1985	1986
MEANS	20.12	19.96	17.40
STANDARD DEVIATION	6.62	5.88	7.66

NUMBER OF ITEMS = 26

Examples

Item 3

$5\overline{)41}$

A 2r4 B 9 C 8r1 D 19

% Correct = 84

Item 9

$9\overline{)99}$

A 11 B 27 C 12r1 D9

% Correct = 78

Item 13

$3\overline{)42}$

A 12 B 14 C 32 D 10r2

% Correct = 64

Item 15

$5\overline{)90}$

A 12 B 14 C 16 D 18

% Correct = 69

Item 19

$7\overline{)76}$

A 17 B 12r2 C 6r6 D 10r6

% Correct = 57

Item 23

$3\overline{)77}$

A 47 B 25r2 C 27 D 22r1

% Correct = 46

Discussion

Average proportion passing an item was 67% compared with 77% in 1985. This has turned out to be one of the more difficult tests in Phase 2. The main problem in the division tasks seems to be in handling regrouping and remainders.

9.5 RELATIONS

(8) Test P2NE2B

Objective: Identify place values.

Results

	1985	1986
MEANS	4.59	4.44
STANDARD DEVIATION	1.00	1.33

NUMBER OF ITEMS = 5

Examples

Item 2

Which numeral is in the tens column?

H	T	U
9	8	7

A 87 B 980 C 8 D 9

% Correct = 91

Item 5

Which numeral is in the units column?

H	T	U
5	4	0

A 40 B 0 C 4 D 540

89 % Correct = 90

Discussion

In this sub-test on place values, average proportion passing an item was 89%, in 1985 it was 92%. It appears that most students knew place values as shown by the high success rates for both 1985 and 1986.

(9) Test P2NE2C

Objective: Write total values.

Results

	1985	1986
MEANS	3.93	3.62
STANDARD DEVIATION	1.18	1.45

NUMBER OF ITEMS = 5

Examples

Item 4

Which number has 2 tens, 6 hundreds and 3 units?

A 263 B 326 C 623 D 362

% Correct = 55

Item 5

Which item has 6 units and 1 hundred?

A 61 B 601 C 16 D 106

% Correct = 63

Discussion

Average proportion successful on an item was 72% compared with 79% in 1985. The success rate was reasonably high but some seem to have a problem with questions put in the form as shown by items 4 and 5.

(10) Test P2NE6A

Objective: Round two digit whole numbers to the nearest 10.

Results

	1985	1986
MEANS	6.24	6.35
STANDARD DEVIATION	1.96	1.83

NUMBER OF ITEMS = 8

Examples

Item 2

Round 92 to the nearest ten.

A 100 B 95 C 90 D 80

% Correct = 84

Item 4

Round 25 to the nearest ten.

A 20 B 30 C 25 D 50

% Correct = 67

Discussion

Average proportion successful on an item was 79% compared with 78% in 1985. The high success rate indicates that the task of rounding to the nearest ten was easy for most children.

(11) Test P2NE6B

Objective: Round three digit whole numbers to the nearest 100.

Results

	1985	1986
MEANS	4.52	4.46
STANDARD DEVIATION	2.13	2.09

NUMBER OF ITEMS = 7

Examples

Item 1

Round 549 to the nearest hundred.

A 400 B 500 C 600 D 550

% Correct = 56

Item 2

Round 827 to the nearest hundred.

A 800 B 830 C 820 D 900

% Correct = 61

Discussion

Average proportion successful on an item was 64% compared with 65% in 1985. The results paralleled those obtained last year. Although most students were able to round numbers to the nearest ten, quite a high proportion had trouble in rounding to the nearest hundred.

CHAPTER TEN

ANALYSIS OF PERFORMANCE ON THE READING TESTS
STAGE 7

ABSTRACT

Three groups of tests were included in reading: comprehension tests using selected passages, tests on dictionary skills and tests in 'reading for different purposes' (RDP).

Performance in reading comprehension was judged on the whole test rather than on parts of it because of the interdependence of objectives in comprehension.

In the case of dictionary skills and 'reading for different purposes', performance on each test reflected performance on an objective.

Competency level at 70% correct was used in the analysis because data collected during 1984 and 1985 confirmed that, in general, the Reading tests were more difficult than the Mathematics tests.

Percentages achieving competence on the tests have been reported. Estimates of sampling error for proportions have been included to enable schools to compare the results of their own students against the Territory results.

For comparing school results with Territory results, graphs have also been provided.

To provide schools with more information on student performance, percentages of students responding correctly to more than 60%, 70%, 80% and 100% of items in each test have been reported.

10.1 THE READING TESTS

The Reading comprehension tests required students to read a given passage and answer a number of multiple-choice questions. These questions were related to objectives in the Reading curriculum. Unlike the Mathematics item pool which consisted of items measuring the core areas, the Reading item pool covered most objectives in the total curriculum for Reading in Years 5 and 7.

The comprehension tests were set on passages selected from materials that were normally encountered in the classroom. Their readability levels had been carefully considered for their suitability by the panel of teachers who had the task of developing the tests.

Twenty five objectives identified in the Reading curriculum for Years 5 and 7 were tested using this group of tests. They covered objectives in both literal and inferential comprehension. However, the main objectives measured by these tests were the following:

(A) Literal Comprehension

Identify main idea
Arrange events in a sequence
Locate specific details and facts
Give meanings of words in context

(B) Inferential Comprehension

Infer from facts given
Draw a conclusion
Identify character traits
Determine cause of event

The next category was a group of four tests on dictionary skills. The main objectives measured by the dictionary tests were:

Use guide words to locate information
Locate main entry and sub-entry words
Select the correct contextual meaning of a multiple meaning word

The third category consisted of tests in 'reading for different purposes', related to the following main objectives:

Locate information from a telephone directory
Follow directions on labels, recipes and signs
Use library reference skills
Interpret information from samples of graphic materials

The cutoff score recommended by the item-writing panels was 80% of items correct for competency although this was not to apply to all tests. The results in 1984 and 1985 confirm that the Reading tests, particularly, the comprehension tests and the tests on dictionary skills, were more difficult than the Mathematics tests. On the basis of data collected, the level of competency for the reading tests was adjusted to 70% correct.

Figure 3 shows overall performance on the tests.

10.2 INTERPRETING STANDARDS

The level of competency or cutoff score was 70% correct. The cutoff score is a measure of standard of performance. In addition to the cutoff score, the proportion of students achieving the cutoff score can be taken as a further measure of standards. However, what constitutes an acceptable proportion is a matter of judgement.

For the purpose of this report, the proportion of 70% of students achieving competence has been employed as a criterion for analysing performance on the objectives. This may be referred to as the 70-70 criterion.

In other words, the 70-70 criterion has been used as the benchmark, that is, 70% of students achieving the competency level fixed at 70% of items correct.

Those objectives in which more than 70% have shown competence are indicated in Table 3B.

10.3 COMPARING SCHOOL RESULTS WITH NORTHERN TERRITORY RESULTS

Table 3A may be used by schools to compare the results of their own students with the results obtained for the Territory.

In this table, estimates of sampling error are shown in the column for 70% correct. These estimates are based on the 95 per cent confidence limits.

For instance, for test C1, in Table 3A, 60.3% of students achieved the competency level fixed at 70% correct. The sampling error for proportions estimated was plus or minus 8% points.

To determine whether differences are significant, any observed difference in this case should be at least 8% point. That is, in a school, the proportion achieving competence on test C1 has to be more than 68% (60.3 + 8.0 = 68.3) for the result to be **significantly better** than the Territory performance.

Conversely, the proportion achieving competence has to be lower than 52% (60.3 - 8.0 = 52.3) for the result to be **significantly worse** than the Territory result.

For comparing school results with the Territory results, the graphs in the Appendix may be used. The 95% confidence limits are indicated in each graph.

Differences required for significance can be read directly from the graph.

10.4 OVERALL ACHIEVEMENT

The results for the comprehension tests involving selected passages and the tests on dictionary skills were generally poorer than those for the tests on reading for different purposes. This observation may be verified in Table 3A or Table 3B.

From these tables, performance across the reading tests for Stage 7 may be summarised as follows:

	TEST GROUP		
	COMPREHENSION	DICTIONARY	READING FOR DIFFERENT PURPOSES
Number of tests where MORE than 70% were competent	3	1	6
Number of tests where LESS than 70% were competent	5	3	2
	8	4	8

10.5 DISCUSSION OF THE RESULTS

An analysis of the results of the tests within each category now follows.

READING COMPREHENSION (SELECTED PASSAGES)

1. Eight tests in reading comprehension based on selected passages were administered during 1986 to Year 7 students.

   Out of 8 tests in this category more than 70% were competent in 3 of these. Between 50% and 70% were competent in the remaining 5.

   The analysis above was based on a cutoff score of 70% correct for competency. Using a single cutoff as a reference point, the interpretation is: a student passes the test if he/she has a score of 70% correct or more; conversely, a student fails the test if he/she has a score less than 70% correct.

A single cutoff might be useful for some purposes but it has a limited value because one does not know by how many marks or items a student is below the competency score. A single cutoff gives what may be described as an on - or - off interpretation of test results. A student is 'on' if he/she passes and 'off' if he/she fails.

2. The information given in Table 3A enables teachers and administrators to interpret the results in terms of four cutoffs or reference points, i.e. 60% correct (or more), 70% correct (or more), 80% correct (or more) and 100% correct.

It is interesting to note that if the competency level was lowered from 70% correct to 60% correct, the data shows that for every test in comprehension with the exception of one, the proportions 'successful' would be more than 70%.

3. The comprehension tests were next ranked in order of their 'difficulty' on the basis of average proportions successful on each item in each test. The results from 1985 and 1986 show quite a high correlation.

1985 Average Proportion (%) Passing an Item	Test Code	Test Code	1986 Average Proportion (%) Passing an Item
85	C7	C2	84
83	C2	C7	82
77	C6	C6	76
75	C8	C8	74
74	C3	C5	73
73	C5	C4	71
72	C4	C1	71
71	C1	C3	64

DICTIONARY SKILLS

1.. As in 1985, the tests on dictionary skills yielded poorer results. More than 40% of students failed to give correct answers to a third of the questions in this category.

Out of 4 tests in this category

More than 70% were competent in one of the tests
Between 50% and 70% were competent in another test
Less than 50% were competent in the remaining two

2. Average proportions passing an item in each test for 1985 and 1986 are shown below.

1985 Average Proportion (%) Passing an Item	Test Code	Test Code	1986 Average Proportion (%) Passing an Item
79	D1	D1	76
68	D2	D4	68
62	D3	D2	65
59	D4	D3	60

TESTS ON READING FOR DIFFERENT PURPOSES

1. The group of tests in 'reading for different purposes' included various reading materials on which questions were asked. The tests measured the following main skills:

Following directions on recipes
Using library reference skills
Interpreting information from various sources such as a street map, breakfast menu, a tourist guide, a sale advertisement, etc.

2. Out of eight tests administered in 1986 in this category more than 70% were competent in six of these. Between 50% and 70% were competent in the remaining two.

3. The following results were obtained for 1985 and 1986 using the same performance indicator: average proportion of students passing an item in each test.

Code	Test	Average Proportion Passing an Item	
		1986	1985
R10	Motel Breakfast Menu	91	83
R14	Encyclopaedia Index	86	84
R9	Katherine Gorge Walks	85	84
R11	Street Map	80	78
R4	Sale at Chemist	78	85
R12	Camping Equipment Retail	72	82
R5	Facts about Poison	67	79
R2	Recipe	63	75

TABLE 3A: PERCENTAGES AT DIFFERENT CUTOFFS: READING STAGE 7

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING CORRECT			
				60%	70%	80%	100%
1	C1	Comprehension (Passage)	19	73.8	60.3 +-8	40.0	6.3
2	C2	Comprehension (Passage)	14	93.7	90.4 +-5	66.6	14.4
3	C3	Comprehension (Passage)	13	68.3	49.6 +-8	33.1	6.5
4	C4	Comprehension (Passage)	20	78.0	60.1 +-8	43.3	1.9
5	C5	Comprehension (Passage)	20	78.8	65.9 +-7	47.6	6.6
6	C6	Comprehension (Passage)	20	86.3	78.5 +-6	59.6	4.8
7	C7	Comprehension (Passage)	10	92.1	86.5 +-5	74.1	21.9
8	C8	Comprehension (Passage)	13	80.0	60.3 +-8	44.0	8.6
9	D1	Dictionary Skills	10	82.9	72.8 +-7	56.9	22.9
10	D2	Dictionary Skills	11	63.2	46.5 +-8	31.9	6.7
11	D3	Dictionary Skills	5	61.6	38.9 +-8	38.9	24.5
12	D4	Dictionary Skills	10	71.2	58.5 +-8	44.1	15.7

TABLE 3A: PERCENTAGES AT DIFFERENT CUTOFFS: READING STAGE 7

NO	CODE	OBJECTIVES	NO OF ITEMS	PROPORTION OBTAINING CORRECT			
				60%	70%	80%	100%
13	R2	RDP Recipe	13	81.9	62.8 +/-8	46.7	12.3
14	R4	RDP Sale at Chemist	9	90.5	77.6 +/-6	59.7	34.8
15	R5	RDP Facts about Poison	11	77.3	66.7 +/-7	54.6	18.3
16	R9	RDP Katherine Gorge Walks	17	87.6	84.8 +/-6	72.8	18.7
17	R10	RDP Motel Breakfast Menu	10	93.3	91.2 +/-5	80.3	28.5
18	R11	RDP Street Map	10	91.1	79.7 +/-6	63.7	11.0
19	R12	RDP Camping Equipment Retail	9	84.4	72.0 +/-7	45.4	19.3
20	R14	RDP Encyclopaedia	5	95.9	85.5 +/-5	85.5	37.9

TABLE 3B: PERCENTAGES ACHIEVING COMPETENCE: READING STAGE 7

NO	CODE	TESTS	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
1	C1	Comprehension (Passage)	19	14	428	60.3*
2	C2	Comprehension (Passage)	14	10	395	90.4
3	C3	Comprehension (Passage)	13	9	369	49.6*
4	C4	Comprehension (Passage)	20	14	268	60.1*
5	C5	Comprehension (Passage)	20	14	410	65.9*
6	C6	Comprehension (Passage)	20	14	497	78.5
7	C7	Comprehension (Passage)	10	7	406	86.5
8	C8	Comprehension (Passage)	13	9	511	60.3*
9	D1	Dictionary Skills	10	7	397	72.8
10	D2	Dictionary Skills	11	8	389	46.5*
11	D3	Dictionary Skills	5	4	229	38.9*
12	D4	Dictionary Skills	10	7	503	58.5*
13	R2	RDP Recipe	13	9	527	62.8*
14	R4	RDP Sale at Chemist	9	6	370	77.6
15	R5	RDP Facts about Poison	11	8	405	66.7*
16	R9	RDP Katherine Gorge Walks	17	12	427	84.8

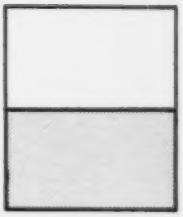
TABLE 3B: PERCENTAGES ACHIEVING COMPETENCE: READING STAGE 7

NO	CODE	TESTS	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
17	R10	RDP Motel Breakfast Menu	10	7	239	91.2
18	R11	RDP Street Map	10	7	482	79.7
19	R12	RDP Camping Equipment Retail	9	6	436	72.0
20	R14	RDP Encyclopaedia Index	5	4	269	85.5

Note: * Indicates less than 70% achieving competence.

C - 1	C - 2	C - 3	C - 4	C - 5
C - 6	C - 7	C - 8	D - 1	D - 2
D - 3	D - 4	R - 2	R - 4	R - 5
R - 9	R - 10	R - 11	R - 12	R - 14

FIG 3: OVERALL PERFORMANCE READING STAGE 7



More than 70% achieved competence

Less than 70% achieved competence

CHAPTER ELEVEN

ANALYSIS OF PERFORMANCE ON THE READING TESTS
STAGE 5

ABSTRACT

Three groups of tests were included in reading: comprehension tests using selected passages, tests of dictionary skills and tests in 'reading for different purposes' (RDP).

Performance in reading comprehension was judged on the whole test rather than on parts of it because of the interdependence of objectives in comprehension.

In the case of dictionary skills and 'reading for different purposes', performance on each test reflected performance on an objective. This objective might be a main objective or a sub-objective.

Competency level at 70% correct was used in the analysis because data collected during 1984 and 1985 confirmed that, in general, the Reading tests were more difficult than the Mathematics tests.

Percentages achieving competence on the tests have been reported. Estimates of sampling error for proportions have been included to enable schools to compare the results of their own students against the Territory results.

For comparing school results with Territory results, graphs have also been provided.

To provide schools with more information on student performance, percentages of students responding correctly to more than 60%, 70%, 80% and 100% of items in each test have been reported.

11.1 THE READING TESTS

The Reading comprehension tests required students to read a given passage and answer a number of multiple-choice questions. These questions were related to objectives in the Reading curriculum. Unlike the Mathematics item pool which consisted of items measuring the core areas, the Reading item pool covered most objectives in the total curriculum for Reading in Years 5 and 7.

The comprehension tests were set on passages selected from materials that were normally encountered in the classroom. Their readability levels have been carefully considered for their suitability by the panel of teachers who had the task of developing the tests.

Twenty five objectives identified in the Reading curriculum for Years 5 and 7 were tested using this group of tests. They covered objectives in both literal and inferential comprehension. However, the main objectives measured by these tests were the following:

(A) Literal Comprehension

Identify main idea
Arrange events in sequence
Locate specific details and facts
Give meanings of words in context

(B) Inferential Comprehension

Infer from facts given
Draw a conclusion
Identify character traits
Determine cause of event

The next category was a group of two tests on dictionary skills. The main objectives measured by the dictionary tests were:

Use guide words to locate information
Select the correct contextual meaning of a word
Order words alphabetically

The third category consisted of tests in 'Reading for different purposes', related to the following main objectives:

Locate information from a telephone directory
Follow directions on labels, recipes and signs
Use library reference skills

The cutoff score recommended by the item-writing panels was 80% of items correct for competency although this was not to apply to all tests. The results in 1984 and 1985 confirmed that the Reading tests, particularly, the comprehension tests and the tests on dictionary skills, were more difficult than the Mathematics tests. On the basis of data collected, the level of competency for the reading tests was adjusted to 70% correct.

Figure 4 shows overall performance on the tests.

11.2 INTERPRETING STANDARDS

The level of competency or cutoff score was 70% correct. The cutoff score is a measure of standard of performance. In addition to the cutoff score, the proportion of students achieving the cutoff score can be taken as a further measure of standards. However, what constitutes an acceptable proportion is a matter of judgement.

For the purpose of this report, the proportion of 70% of students achieving competence has been employed as a criterion for analysing performance on the objectives. This may be referred to as the 70-70 criterion.

In order words, the 70-70 criterion has been used as the benchmark, that is, 70% of students achieving competence fixed at 70% of items correct.

Those objectives in which more than 70% have shown competence are indicated in Table 4B.

11.3 COMPARING SCHOOL RESULTS WITH NORTHERN TERRITORY RESULTS

Table 4A may be used by schools to compare the results of their own students with the results obtained for the Territory.

In this table, estimates of sampling error are shown in the column for 70% correct. These estimates are based on the 95 per cent confidence limits.

For instance, for test C2, in Table 4A, 62.7% of students achieved the competency level fixed at 70% correct. The sampling error for proportions estimated was plus or minus 8% points.

To determine whether differences are significant, any observed difference in this case should be at least 8% points. That is, in a school, the proportion achieving competence on test C2 has to be more than 71% (62.7 + 8.0) for the result to be **significantly better** than the Territory performance.

Conversely, the proportion achieving competence has to be lower than 55% (62.7 - 8.0) for the result to be **significantly worse** than the Territory result.

For comparing school results with the Territory results, the graphs in the Appendix may be used. The 95% confidence limits are indicated in each graph.

Differences required for significance can be read directly from the graph.

11.4 OVERALL ACHIEVEMENT

The results for the comprehension tests involving selected passages and the tests on dictionary skills were generally poorer than those for the tests on reading for different purposes. A similar pattern was found in 1985. This observation may be verified in Table 4A or Table 4B.

From these tables, performance across the reading tests for Stage 5 may be summarised as follows:

	TEST GROUP		
	COMPREHENSION	DICTIONARY	READING FOR DIFF PURPOSES
Number of tests where MORE than 70% were competent	1	0	5
Number of tests where LESS than 70% were competent	6	2	3
	7	2	8

11.5 DISCUSSION OF THE RESULTS

An analysis of the results of the tests within each category now follows.

READING COMPREHENSION (SELECTED PASSAGES)

1. Seven tests in reading comprehension based on selected passages were administered during 1986 to Year 5 students.

   Out of 7 tests in this category

   More than 70% were competent in one test
   Between 50% and 70% were competent in three tests
   Less than 50% were competent in the remaining three tests

   The analysis above was based on a cutoff score of 70% correct for competency. Using a single cutoff as a reference point, a student passes the test if he/she has a score of 70% correct, or more; conversely, a student fails the test if he/she has a score less than 70% correct.

A single cutoff might be useful for some purposes but it has a limited value because one does not know by how many marks or items a student is below the competency score. A single cutoff gives what may be described as an on - or - off interpretation of test results. A student is 'on' if he/she passes and 'off' if he/she fails.

2. The information given in Table 4A enables teachers and administrators to interpret the results in terms of four cutoffs or reference points, i.e. 60% correct (or more), 70% correct (or more), 80% correct (or more) and 100 correct.

3. The comprehension tests were next ranked in order of their 'difficulty' on the basis of average proportions successful on each item in each test.

Test Code	Average Proportion Passing An Item (%)	
	1986	1985
C6	78	78
C2	72	72
C4	70	69
C7	64	64
C5	63	70
C3	63	69
C1	58	61

4. For both years, C6 was the easiest test and C1 the most difficult.

DICTIONARY SKILLS

The table below shows average proportions passing an item on the two tests in this category.

Test Code	Average Proportion Passing An Item (%)	
	1986	1985
D1	61	63
D2	42	43

As in 1985, the more difficult test was D2 which required students to use guide words in locating information.

TESTS ON READING FOR DIFFERENT PURPOSES

1. The group of tests in 'reading for different purposes' included various reading materials on which questions were asked. The tests measured the following main skills:

   Locating information from a telephone directory
   Following directions on recipes, safety
   Using library reference skills
   Interpreting information from various sources such as a street map, a poster, a concert program

2. Out of eight tests administered in this category more than 70% were competent in five of these. Between 50% and 70% were competent in two of these and less than 50% were competent in the remaining one.

3. On the basis of average proportions passing an item in each of the tests, the following was the order in terms of average difficulty:

Test Code	Average Proportion Passing An Item (%)	
	1986	1985
R3: Emergency Calls	85	87
R2: Street Map	80	76
R10: School Concert Program	79	73
R8: Street Map	75	71
R7: Contents of Book	73	76
R12: TV Program	70	73
R1: RDP Recipe	63	-

Generally, performance was very satisfactory. With the exception of one test, average proportion passing an item in the tests in this group was more than 70%.

TABLE 4A: PERCENTAGES AT DIFFERENT CUTOFFS: READING STAGE 5

NO	CODE	TESTS	NO OF ITEMS	PROPORTION OBTAINING CORRECT			
				60%	70%	80%	100%
1	C1	Comprehension (Passage)	12	43.6	31.1 +-7	18.8	2.3
2	C2	Comprehension (Passage)	15	78.0	62.7 +-8	51.5	8.9
3	C3	Comprehension (Passage)	12	55.2	41.9 +-8	31.5	4.1
4	C4	Comprehension (Passage)	10	80.7	69.1 +-7	51.9	1.4
5	C5	Comprehension (Passage)	10	65.6	55.6 +-8	38.0	6.0
6	C6	Comprehension (Passage)	10	85.0	78.6 +-6	69.6	27.4
7	C7	Comprehension (Passage)	11	64.5	48.4 +-8	31.8	6.3
8	D1	Dictionary Skills	8	58.7	41.6 +-8	24.6	7.8
9	D2	Dictionary Skills	5	32.8	17.9 +-6	17.9	9.2
10	R1	RDP Recipe	14	57.9	45.3 +-8	24.9	5.9
11	R2	RDP Street Map	11	90.2	80.7 +-6	68.0	15.3
12	R3	RDP Emergency Calls	11	89.0	84.9 +-6	77.7	37.9

TABLE 4A: PERCENTAGES AT DIFFERENT CUTOFFS: READING STAGE 5

NO	CODE	TESTS	NO OF ITEMS	PROPORTION OBTAINING CORRECT			
				60%	70%	80%	100%
13	R7	RDP Contents of Book	10	81.5	70.0 +7	53.3	13.7
14	R8	RDP Street Map	10	85.2	77.9 +-6	60.1	11.7
15	RD10	RDP School Concert Program	11	85.7	78.3 +-6	60.0	19.0
16	R11	RDP Child Vacation Care Program	12	77.4	68.9 +-7	53.9	0.2
17	R12	RDP TV Program	11	72.0	60.7 +-8	46.7	7.2

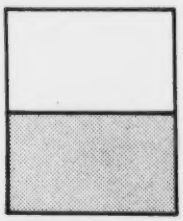
TABLE 4B: PERCENTAGES ACHIEVING COMPETENCE: READING STAGE 5

NO	CODE	TESTS	NO OF ITEMS	NO OF ITEMS REQUIRED	NO OF STUDENTS	PERCENTAGE ACHIEVING COMPETENCE
1	C1	Comprehension (Passage)	12	9	872+	31.1*
2	C2	Comprehension (Passage)	15	11	573	62.7*
3	C3	Comprehension (Passage)	12	9	830+	41.9*
4	C4	Comprehension (Passage)	10	7	285	69.1*
5	C5	Comprehension (Passage)	10	7	550	55.6*
6	C6	Comprehension (Passage)	10	7	299	78.6
7	C7	Comprehension (Passage)	11	8	349	48.4*
8	D1	Dictionary Skills	8	6	859+	41.6*
9	D2	Dictionary Skills	5	4	955+	17.9
10	R1	RDP Recipe	14	10	558	45.3*
11	R2	RDP Street Map	11	8	347	80.7
12	R3	RDP Emergency Calls	11	8	601	89.9
13	R7	RDP Contents of Book	10	7	437	70.0
14	R8	RDP Street Map	10	7	589	77.9
15	R10	RDP School Concert Program	11	8	300	78.3
16	R11	RDP Child Vacation Care Program	12	9	562	68.9*
17	R12	RDP TV Program	11	8	567	60.7*

(Note: * indicates less than 70% achieving competence
+ two groups given the same test)

C - 1	C - 2	C - 3	C - 4	C - 5
C - 6	C - 7	D - 1	D - 2	R - 1
R - 2	R - 3	R - 7	R - 8	R - 10
			R - 11	R - 12

FIG 4: OVERALL PERFORMANCE READING STAGE 5



More than 70% achieved competence

Less than 70% achieved competence

CHAPTER TWELVE

COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE READING STAGES 5 AND 7

ABSTRACT

In this chapter, performance in 1986 was compared with that in 1985 on each of the tests in Reading. The performance indicator used was the percentage of students achieving competence on each test. The results showed that there were only a few tests in which a significant improvement or a significant decline was observed. In the majority of tests, changes in performance were not significant. There was no strong evidence to suggest an overall improvement or decline in any of the areas in reading in 1986 for both Years 5 and 7.

12.1 CHANGES IN PERFORMANCE

The performance indicator used in comparing the results in 1985 with those in 1986 was the percentage of students achieving competence.

Table 3C and Table 4C show percentages achieving competence on each test for 1985 and 1986. The change in performance is given by the difference between the results of the two years. A plus sign (+) indicates that the change was an improvement, a negative sign (-) shows that there was a decline.

Changes in performance are statistically significant if there was a 95% probability that the differences between the results for the two years were real and could not have come about by chance. Changes in performance that are statistically significant are shown with an asterisk.

12.2 READING TESTS - STAGE 7

Table 3C shows changes in performance between 1985 and 1986 on the reading tests for Stage 7. Twenty tests, eight in reading comprehension, four on dictionary skills and eight in reading for different purposes were administered in both years.

12.3 COMPREHENSION OF PROSE PASSAGE

Table 3D provides a summary of changes in performance for each group of tests in Stage 7.

Of the eight tests in comprehension using selected passages, the results showed marginal changes in seven and a significant decline in one.

Since the results from seven out of eight tests in this group indicated only marginal changes, there was no evidence to suggest an improvement or a decline in the ability of Year 7 children in comprehension of prose passages.

12.4 DICTIONARY SKILLS

In the group of tests on dictionary skills, a significant improvement was observed from the results of one test and a significant decline from the results of two of the tests. A marginal decline was noted in the results from the remaining test.

The results seem to suggest that there was a decline in dictionary skills in 1986.

12.5 READING FOR DIFFERENT PURPOSES

Eight tests in reading for different purposes were administered in both 1985 and 1986. The results showed marginal changes from the results of seven out of eight tests but a significant decline from the results of the remaining one.

As stated earlier, changes that were marginal could have come about by chance. Hence, there was no evidence to suggest that basic reading skills involved in comprehension of materials frequently encountered such as an advertisement, a menu, street map, an encyclopaedia index have improved or declined in 1986.

12.6 READING TESTS : STAGE 5

Table 4C shows changes in performance between 1985 and 1986 on the reading tests for Stage 5. Seventeen tests, seven in reading comprehension, two on dictionary skills and eight in reading for different purposes were administered in both years.

12.7 COMPREHENSION OF PROSE PASSAGE

Table 4D gives a summary of changes in performance for each group of tests in Stage 5.

Seven of these were tests in comprehension where students had to answer questions after reading a given prose passage. The results from five of these tests showed marginal improvement or decline which could have come about by chance. However, the results from two tests indicated a significant decline compared with results in 1985.

12.8 DICTIONARY SKILLS

There were two tests on dictionary skills. The results showed a drop in performance in both but these were only marginal and not significant. There was no evidence to suggest an improvement or a decline in the use of dictionary skills of Year 5 children.

12.9 READING FOR DIFFERENT PURPOSES

In this category, there were eight tests covering materials such as a recipe, street map, contents of a book, concert program, poster, TV program and telephone directory. Children were asked questions requiring them to extract information from materials that they would frequently encounter.

As given in Tables 4C and 4D, of eight tests in this group, the results from two indicated a significant improvement and results from another two showed a significant decline. However, the results from four of the eight tests showed no significant changes. Overall, it appears that the levels of achievement of Year 5 children in reading for different purposes have neither improved nor declined in 1986.

TABLE 3C: CHANGES IN PERFORMANCE: READING STAGE 7

NO	CODE	TESTS	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
1	C1	Comprehension (Passage)	19	56.3	60.3	+4.0
2	C2	Comprehension (Passage)	14	87.5	90.4	+2.9
3	C3	Comprehension (Passage)	13	58.2	49.6	-8.6*
4	C4	Comprehension (Passage)	20	65.4	60.1	-5.3
5	C5	Comprehension (Passage)	20	67.1	65.9	-1.2
6	C6	Comprehension (Passage)	20	77.6	78.5	+0.9
7	C7	Comprehension (Passage)	10	90.3	86.5	-3.8
8	C8	Comprehension (Passage)	13	60.3	60.3	0.0
9	D1	Dictionary Skills	10	76.9	72.8	-4.1
10	D2	Dictionary Skills	11	57.9	46.5	-11.4*
11	D3	Dictionary Skills	5	47.2	38.9	-8.3*
12	D4	Dictionary Skills	10	45.4	58.5	+13.1*
13	R2	RDP Recipe	13	62.5	62.8	+0.3
14	R4	RDP Sale at Chemist	9	80.8	77.6	-3.2
15	R5	RDP Facts about Poison	11	77.4	66.7	-10.4*
16	R9	RDP Katherine Gorge Walks	17	87.3	84.8	-2.5
17	R10	RDP Motel Breakfast Menu	10	86.6	91.2	+4.6
18	R11	RDP Street Map	10	80.9	79.7	-1.2
19	R12	RDP Camping Equipment Retail	9	77.3	72.0	-5.3
20	R14	RDP Encyclopaedia Index	5	82.2	85.5	+3.3

TABLE 4C: CHANGES IN PERFORMANCE: READING STAGE 5

NO	CODE	TESTS	NO OF ITEMS	PERCENTAGE ACHIEVING COMPETENCE		CHANGE IN PERFORMANCE
				1985	1986	
1	C1	Comprehension (Passage)	12	35.8	31.1	-4.7
2	C2	Comprehension (Passage)	15	62.4	62.7	+0.3
3	C3	Comprehension (Passage)	12	53.9	41.9	-12.0*
4	C4	Comprehension (Passage)	10	64.4	69.1	+4.7
5	C5	Comprehension (Passage)	10	67.1	55.6	-8.5*
6	C6	Comprehension (Passage)	10	80.9	78.6	-2.3
7	C7	Comprehension (Passage)	11	47.5	48.4	+0.9
8	D1	Dictionary Skills	8	48.7	41.6	-7.1
9	D2	Dictionary Skills	5	20.0	17.9	-2.1
10	R1	RDP Recipe	14	54.2	45.3	-8.9*
11	R2	RDP Street Map	11	74.0	80.7	+6.7
12	R3	RDP Emergency Calls	11	88.0	89.9	+1.9
13	R7	RDP Contents of Book	10	75.2	70.0	-5.2
14	R8	RDP Street Map	10	68.6	77.9	+9.3*
15	R10	RDP School Concert Program	11	68.6	78.3	+9.7*
16	R11	RDP Child Vacation Care Program	12	84.0	68.9	-15.1*
17	R12	RDP TV Program	11	68.5	60.7	-7.8

TABLE 3D: READING TESTS: YEAR 7

CHANGES IN PERFORMANCE : SUMMARY

	Comprehension of of Passage	Dictionary Skills	Reading for Different Purposes
Number of Tests Administered	8	4	8
Number of tests in which significant improvement was shown	0	1	0
Number of tests in which significant decline was shown	1	2	1
Number of tests in which changes were not significant	7	1	7

TABLE 4D: READING TESTS: YEAR 5

CHANGES IN PERFORMANCE : SUMMARY

	Comprehension of of Passage	Dictionary Skills	Reading for Different Purposes
Number of Tests Administered	7	2	8
Number of tests in which **significant improvement** was shown	0	0	2
Number of tests in which **significant decline** was shown	2	0	2
Number of tests in which changes were **not significant**	5	2	4

CHAPTER THIRTEEN

SUMMARY AND CONCLUSION

13.1 THE PRIMARY ASSESSMENT PROGRAM

This is the third Departmental report on the performance of students in the Primary Assessment Program since it was introduced in all urban schools in 1984. Currently, the program has three components:

(1) a collection of tests or item pools in reading and mathematics for Years 5 and 7;

(2) a writing moderation package which includes a description of standard procedures for moderation of children's written compositions; and

(3) a collection of practical strategies for assessment of reading in areas where pencil and paper tests would be inappropriate or less meaningful.

The main aims of the program were to (1) assess student achievement in the essential skills and understandings identified in the English and mathematics core curricula, and (2) identify broad areas of strength and weakness in student performance in the core components in order that remedial provisions for student deficiencies can be made at both the school and system levels. The collections of assessment materials were designed to assist teachers in their instruction. As the materials are criterion-referenced in nature, they measure performance against the core objectives and are not intended to be used for comparing performance between individuals or between groups of students.

13.2 A WORD ON PERFORMANCE STANDARDS

In recent years, a lot of literature has emerged on methods of setting standards for criterion-referenced tests. Standard-setting involves deciding upon the level of student achievement that constitutes an acceptable performance. The fact remains that all standard-setting methods are

arbitrary. They are arbitrary because they involve judgements of one kind or another.

In the Northern Territory Primary Assessment Program, for instance, the standard for acceptable performance on the tests in the item pools is 80% correct. This means that in a 10-item test, the cutoff score is 8 items correct. However, the 80% correct level has not been applied universally to all the tests in the item collections. Some tests have lower cutoff scores; others have higher. In the analysis of the results, the cutoff scores for the reading and mathematics tests were 70% correct and 80% correct respectively.

13.1 COLLECTION OF DATA

For the collection of data, urban schools in the Northern Territory were assigned into four groups. Each group included large, medium and small schools (based on student enrolment). Sampling was done by region and school size.

Each school was asked to administer a sample of tests in reading and mathematics at each of the two year levels. The tests had been selected from the item pools and then randomly assigned to each of the four groups of schools. It might be assumed that most schools carried out their testing during the latter part of the second semester in 1986 as it was during this period that most schools returned their students' answer sheets.

To protect the confidentiality of information collected, each school was given a school code known only to the project coordinator. Each student in Year 5 and Year 7 was also given a student number by the school. The school code and student number were used to identify each answer sheet as the data were being keyed into the computer for processing and analysis.

13.4 ANALYSIS OF RESULTS

In the main report, the results have been analysed and reported as follows:

(1) a set of tables shows percentages of students responding correctly to 70%, 80%, 90% and 100% of items. In reading, the 60% cutoff has also been included as a reference point;

(2) another set of tables shows percentages of students achieving competence on each test/objective;

(3) graphs are provided that show percentages of students achieving competence on each test. Estimates of sampling error are included in the graphs. Individual school results may be compared against the Territory results using the tables or the graphs;

(4) a diagram illustrates performance in each subject area for each year level;

(5) comparisons are made between the results for 1985 and 1986 on each test/objective and changes in performance are discussed;

(6) the results of some tests in mathematics have been analysed further to give the reader an idea of the test content. Several items have been reproduced and the proportions successful on each of these reported.

13.5 ANALYSIS OF PERFORMANCE ON THE MATHEMATICS TESTS: PHASE 3

13.51 PERFORMANCE ON OBJECTIVES

In Phase 3, twenty eight objectives were tested and the following were the results:

On 16 objectives, more than 70% were competent.
On 7 objectives, between 50% and 70% were competent.
On 5 objectives, less than 50% were competent.

As discussed in the main report, the proportion '70% of students achieving competence' has been used as an arbitrary benchmark. Using this criterion, the following is a list of the main areas in which more than 70% of Year 7 children have shown competence:

Identify position using ordered pairs.
Know relationship between mm, cm, m and km.
Find and compare areas by counting.

Know 1 L = 1 000 mL
Identify symbols for length.
Add 4-digit numbers.
Solve word problems using addition, multiplication.
Multiply and divide numbers involving decimals.
Show order of fractions (like denominators).

Areas where less than 50% have shown competence would be considered as areas of weakness. The following lists the specific objectives where less than 50% of students have attained the cutoff score:

Measure angles between 0 deg and 180 deg.
Convert units of length.
Convert g into kg and vice versa.
Find interval between two given times.
Identify symbols for volume.

Quite a high proportion of children in Year 7 seem to have a problem in converting one unit of length to another where decimal computation was involved. For instance, 638 cm = _____ m A 0.638 m B 63.8 m C 6 380 m D 6.38 m.

In another test which required changing one unit of weight into another, most seem unable to give the correct answer as in the example: 8 kg 75 g is the same as A 0.875 g B 8 075 g C 8 705 g D 8 750 g.

Both these areas have had consistently poor results in the last three years. Another area of difficulty was in calculating the time interval between two stated times, for instance, between 2:20 pm and 4:00 pm. Most children did not have any trouble in identifying the correct symbols for units of length and weight, e.g. kilometre = km, or kilogram = kg. In the test on the units for volume, most children gave the correct symbol for litre; but a high proportion seem to be confused with the symbol for millilitres which is mL.

13.52 COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE: MATHEMATICS PHASE 3

The results of fifteen tests common to both 1985 and 1986 administrations were compared. The findings show a significant improvement on five objectives, a significant decline on five and no change on the other five objectives.

Significant improvement was noted in these areas: division (decimals included) with regrouping and remainders, using ordered pairs to identify positions on a grid and the conversion 1 L = 1 00 mL. Significant decline was noted mainly in these areas: conversion of units g into kg and vice versa and conversion involving units of length.

Hence, it appears that performance was rather weak in areas involving operations with decimals as shown by the results in the tests on the conversion of units. For instance, the change from centimetres to metres as in the example, 7 cm = ____ m A 70 m B 0.7 m C 0.07 m D 700 m.

A similar problem was seen in the conversion involving units of weight as in the following example, 7 kg 362 g is the same as A 7.362 g B 73.62 g C 736.2 g D 7 362 g.

Overall, however, on the majority of areas tested in Phase 3 at Year 7, the observation was either a significant improvement or no significant change in performance between 1985 and 1986.

In the main report, there are two chapters dealing with the results of individual questions in the mathematics tests for Years 5 and 7 (see chapters six and nine).

13.6 ANALYSIS OF PERFORMANCE ON THE MATHEMATICS TESTS: PHASE 2

13.61 PERFORMANCE ON OBJECTIVES

The following were the results on the thirty three objectives tested in Phase 2 in 1986:

On 20 objectives, more than 70% were competent.
On 9 objectives, between 50% and 70% were competent.
On 4 objectives, less than 50% were competent.

The following were the main areas where more than 70% of Year 5 children had demonstrated competence:

Compare areas.
Read a calendar.
Read time (digital).
Add and subtract whole numbers.
Multiply whole numbers.
Divide money without remainders.
Identify place values.
Order relations between numbers.
Read bar graphs.

On the other hand, there were a few specific objectives where less than 50% have shown competence and these would be considered as areas of weakness:

Use ordered pairs to find positions on a grid.
Read time (colloquial).
Divide two digit numbers by single digits, with and without remainders.
Round to nearest hundred.

In Phase 2, the main areas of weakness seem to be the following: reading time shown on a clockface particularly when the minute hand was not on the hour, half hour or quarter hour; division tasks that involved regrouping and remainders as in this example, 7)76 = _____ A 17 B 12 r2 C 6 r6 D 10 r6.

13.62 COMPARISON BETWEEN 1985 AND 1986 PERFORMANCE: MATHEMATICS PHASE 2

There were twenty four tests common to both the 1985 and 1986 administrations. The results from two showed a significant improvement and the results from nine indicated no significant change. However, the results on the other thirteen objectives showed a significant decline. The main areas that showed a significant drop in performance in 1986 were in addition involving decimals, multiplication and division and the use of ordered pairs to locate position in space.

It might be pointed out that certain subtraction and division tasks posed a problem to quite a high proportion of children. These were subtraction involving three digit numbers with decimals and division of two digit numbers by single digits, with and without remainders. In subtraction, children had difficulty with regrouping; for instance, 75.2 - 30.6 = _____. In division, the main problem seems to be in handling regrouping and remainders as in the following example, 3)77 = _____.

A decline in performance from an already high level in 1985 might not warrant a great deal of concern. For instance, in one area 'reading a calendar', in 1985, 87.0% of Year 5 students were successful but in 1986, 77.4% were successful. This was the case for seven out of the nine areas in which a significant decline was noted.

Of somewhat greater concern would be those areas where there had been rather poor performance in the last two or three years. At the Year 5 level, these would be reading time (colloquial), subtraction involving decimals, division tasks involving regrouping and remainders and rounding to the nearest 10 or 100.

Whilst performance on objectives might be helpful in identifying areas of strength and weakness in the curriculum, the report has included a discussion on performance on individual items for some selected tests.

13.7 ANALYSIS OF PERFORMANCE ON THE READING TESTS

13.71 DATA ANALYSIS

Three groups of tests were included in reading: comprehension tests using selected passages, dictionary skills and tests in reading for different purposes. Performance in reading comprehension was judged on the whole test rather than on parts of it because of the interdependence of objectives in comprehension. In the case of dictionary skills and reading for different purposes, performance on each test reflected performance on an objective.

Competency level was fixed at 70% correct instead of 80% correct because data collected during the last three years confirmed that, in general, the reading tests were more difficult than the mathematics tests. Percentages achieving competence on the tests have been reported. Estimates of sampling error for proportions have also been included to enable schools to compare the results of their own students against the Territory results.

Graphs showing percentages successful on each test are shown in the appendix. These graphs include sampling errors.

To provide schools with more information on student performance, percentages responding correctly to more than 60%, 70%, 80% and 100% of items in each test have been tabulated.

13.8 AREAS TESTED IN READING

The reading comprehension tests required students to read a given passage and answer a number of multiple-choice questions. These questions were related to the objectives in the core curriculum. The comprehension tests were set on passages selected from materials normally encountered in the classroom. Twenty five objectives identified in the reading curriculum for Years 5 and 7 were tested and they included the following:

(a) Literal comprehension: identify main idea; arrange events in a sequence; locate specific details and facts; and give meanings of words in context.

(b) Inferential comprehension: infer from facts given; draw a conclusion; identify character traits; and determine cause of event.

The second category was a group of tests on dictionary skills. The main objectives measured by these tests were: use guide words to locate information; locate main entry and sub-entry words and select the correct contextual meaning of a multiple meaning word.

The third category consisted of a series of tests in reading for different purposes. The objectives tested by this category included: locate information from a telephone directory; follow directions on labels, recipes; use library reference skills; and interpret information from samples of everyday materials such as a street map, a guide on walking tracks at Katherine Gorge, a TV program, a concert program or an advertisement. This group of tests was intended to give students the opportunity to respond to everyday situations which demand the mastery of basic literacy skills.

13.9 SUMMARY OF RESULTS IN THE READING TESTS : STAGE 7

Below is a summary of the results obtained for the three categories of reading tests administered to Year 7 children in 1986:

	TEST GROUP		
	Comprehension	Dictionary	Reading for Diff Purposes
Number of tests where more than 70% were competent	3	1	6
Number of tests where less than 70% were competent	5	3	2
COLUMN TOTALS	8	4	8

There were eight tests in comprehension, four on dictionary and eight in reading for different purposes. They made up the tests measuring all the major objectives in the core curriculum.

The results show that more than 70% of students achieved the cutoff scores on three out of the eight tests in comprehension. It is interesting to note that if the cutoff score was lowered from 70% correct to 60% correct, more than 70% of students would have been 'successful' in all the tests except one in this group. Another indicator used was 'average proportion successful on an item'. Average proportions successful on an item ranged from 64% on the hardest test to 84% on the easiest test. The results from the previous year were very similar.

As in 1985, the tests on dictionary skills yielded poorer results. More than 40% of students failed to give correct answers to a third of the questions contained in this category.

Out of the eight tests in the third category, more than 70% were competent in six of these. Between 50% and 70% were competent in the remaining two. Average proportions successful on an item ranged from 63% on the hardest test to 91% on the easiest test. The results on the tests in basic literacy skills required in reading for different purposes were very pleasing.

13.10 SUMMARY OF CHANGES IN PERFORMANCE IN READING : STAGE 7

The table below gives a summary of changes in performance between 1985 and 1986 on the reading tests for Year 7:

	TEST GROUP		
	Comprehension	Dictionary	Reading for Diff Purposes
Number of tests where significant improvement was noted	0	1	0
Number of tests where significant decline was noted	1	2	1
Number of tests where no significant changes were noted	7	1	7
COLUMN TOTALS	8	4	8

Twenty tests, eight in reading comprehension, four on dictionary skills and eight in reading for different purposes were administered in both 1985 and 1986.

Of the eight tests in comprehension, the results showed marginal changes in seven and a significant decline in one. Since the results from seven out of eight tests in this category indicated only marginal changes between 1985 and 1986, there was no evidence to suggest either an improvement or a decline in the ability of Year 7 children to comprehend prose passages presented.

In dictionary skills, the results seem to suggest a bit of a decline. As stated in previous reports on this program, the tests on dictionary skills were less valid than the other tests in reading. Dictionary skills would be more meaningfully tested in a practical situation.

In the final category of reading tests, the results showed marginal changes in seven out of the eight but a significant decline in one. Again, there was no evidence to suggest either an improvement or a decline in basic literacy skills demanded in everyday situations where children were required to perform the following tasks. These were: extract and interpret information from various sources such as a menu or recipe, an advertisement, a street map, or a tourist brochure.

13.11 SUMMARY OF RESULTS IN THE READING TESTS : STAGE 5

The results on the comprehension tests and on dictionary skills were somewhat poorer than those on reading for different purposes as the summary table below shows:

	TEST GROUP		
	Comprehension	Dictionary	Reading for Diff Purposes
Number of tests where more than 70% were competent	1	0	5
Number of tests where less than 70% were competent	6	2	3
COLUMN TOTALS	7	2	8

Seven tests in reading comprehension were administered during 1986 to Year 5 students. Out of seven tests in this category, there was one test in which more than 70% of children were competent, there were three where between 50% and 70% were competent and three others where less than 50% were competent. Average proportions passing an item ranged from 58% on the hardest test to 78% on the easiest test. The results, in general, seem to suggest that quite a fair proportion of children in Year 5 were not successful in the comprehension tests.

Performance on dictionary skills was very similar to that of the previous year. There appears to be some weakness in this area.

Performance in the last category demanding basic literacy skills was much better. The results from eight tests in this group showed that more than 70% of children were competent in five of them. Between 50% and 70% were competent in two others and less than 50% were competent in the remaining one. Average proportions successful on an item ranged from 63% on the hardest test to 85% on the easiest. There was enough evidence to suggest that most Year 5 children were able to comprehend and extract the relevant information from materials that they encountered frequently.

13.2 SUMMARY OF CHANGES IN PERFORMANCE IN READING : STAGE 5

The table below summarises the changes in performance between 1985 and 1986 on the reading tests administered to Year 5 children:

	TEST GROUP		
	Comprehension	Dictionary	Reading for Diff Purposes
Number of tests where significant improvement was noted	0	0	2
Number of tests where significant decline was noted	2	0	2
Number of tests where changes were not significant	5	2	4
COLUMN TOTALS	7	2	8

Out of seven tests in comprehension, the results from five showed marginal improvement or marginal decline; these do not indicate any significant change. However, the results from two tests showed a significant decline. There was insufficient evidence to suggest any improvement or decline in comprehension at the Year 5 level.

The results from the dictionary tests showed no significant change between the two years.

In the third category, there were eight tests on comprehension of materials commonly encountered. The results from four of the eight tests showed no significant changes. There was a significant improvement on two but a significant decline on two others. Overall, it appears that the basic literacy skills of Year 5 children, as shown from the results on these tests, have neither improved nor declined in 1986.

13.13 CONCLUSION

It needs to be emphasised that the Primary Assessment Program encompasses a wider range of assessment measures than just pencil and paper tests. Moderation of children's writing and application of other informal strategies for assessing reading have already been put in place in Years 5 and 7 in urban schools. Collections of test items with some known statistical characteristics derived from populations in which they have been used are complementary to the other measures employed. Together, they seem to make up a balanced package that teachers could tap into for assessing the core areas in English and mathematics.

The performance of Year 5 and Year 7 children in the core areas of reading and mathematics was the focus of this report. The 'pass score' employed was 80% of items correct for the mathematics tests and 70% correct for the reading tests. The discrepancy was explained by the content differences in the tests. Overall, the reading test items, particularly in the comprehension tests, appear to be more difficult than the mathematics test items.

Reference points of 60, 70, 80 and 100 percent of items correct were used to illustrate success rates associated with several possible cutoff scores.

A number of factors must be considered in analysing the performance of students in mathematics. The tests in mathematics were set at the level of the core. Most of the areas tested would have been covered in earlier years. The tests were mainly of the computational type; students were asked to perform routine computational and measurement tasks.

The results in mathematics at both year levels showed clearly that the majority of children have mastered most of the core areas tested. There were only a few areas in which performance appeared to be unsatisfactory. At Year 7, problems dealing with conversion of units, for instance, changing x cm into m or y kg into g appeared to be particularly difficult. Operations involving decimals posed some problems to quite a high proportion of children. Measurement of angles and calculating the interval between two stated times also emerged as the more difficult areas. Once again, it needs to be stressed that these were confined to isolated items within the tests.

At the Year 5 level, the following areas appeared to be particularly difficult: reading time (colloquial), subtraction involving decimals, division involving regrouping and remainders and rounding.

Overall, in mathematics, when performance in 1985 was compared with that in 1986, there was no evidence of any significant change at the Year 7 level. At Year 5, a significant decline was noted in some isolated areas but in most cases, the drop in performance was in areas which had high levels of performance in 1985.

To give the reader some idea of the item content, some of the test items have been reproduced and performance on these discussed. It has not been possible to do the same for reading because it would have meant reproduction of the whole reading/stimulus materials and test items.

The results from the reading tests showed that the achievement levels of Year 7 children appeared to be higher than those of Year 5 children particularly in comprehension of prose passages (Year 7 and Year 5 children took different sets of comprehension tests). This might be due, in part, to the fact that the passages set for Year 5 were perhaps difficult for this age group.

The comprehension tests appeared to be more difficult than the tests in reading for different purposes. This was to be expected because of content differences, that is, the prose passages given for the comprehension tests were more difficult than the materials presented for reading for different purposes.

Performance on the comprehension tests was generally satisfactory at Year 7, but at Year 5, performance levels were somewhat poorer. On the tests in reading for different purposes, performance was at a reasonably high level for both age groups. The tests on dictionary skills yielded rather poor results in both groups.

In general terms, on the reading tests, at both the Year 5 and Year 7 levels, comparison of performance between 1985 and 1986 showed that overall there was neither a significant improvement nor a significant decline in all three categories of tests used.

Once again, it is necessary to remind ourselves that the Primary Assessment Program is concerned with only the most basic skills in mathematics and reading. This is particularly true for mathematics where the core knowledge and skills have probably been acquired two years before the testing. The tests used in the program were not demanding and were not expected to stretch the ability of the more able children. These tests were designed to monitor standards on the basic skills. Interpretations of changes in standards should therefore be treated with caution. Changes in performance have been reported and discussed in relation to specific areas. It might be more helpful to interpret changes in performance in respect of specific areas rather than in the broader context of standards generally.

FIG 1.1 SPACE

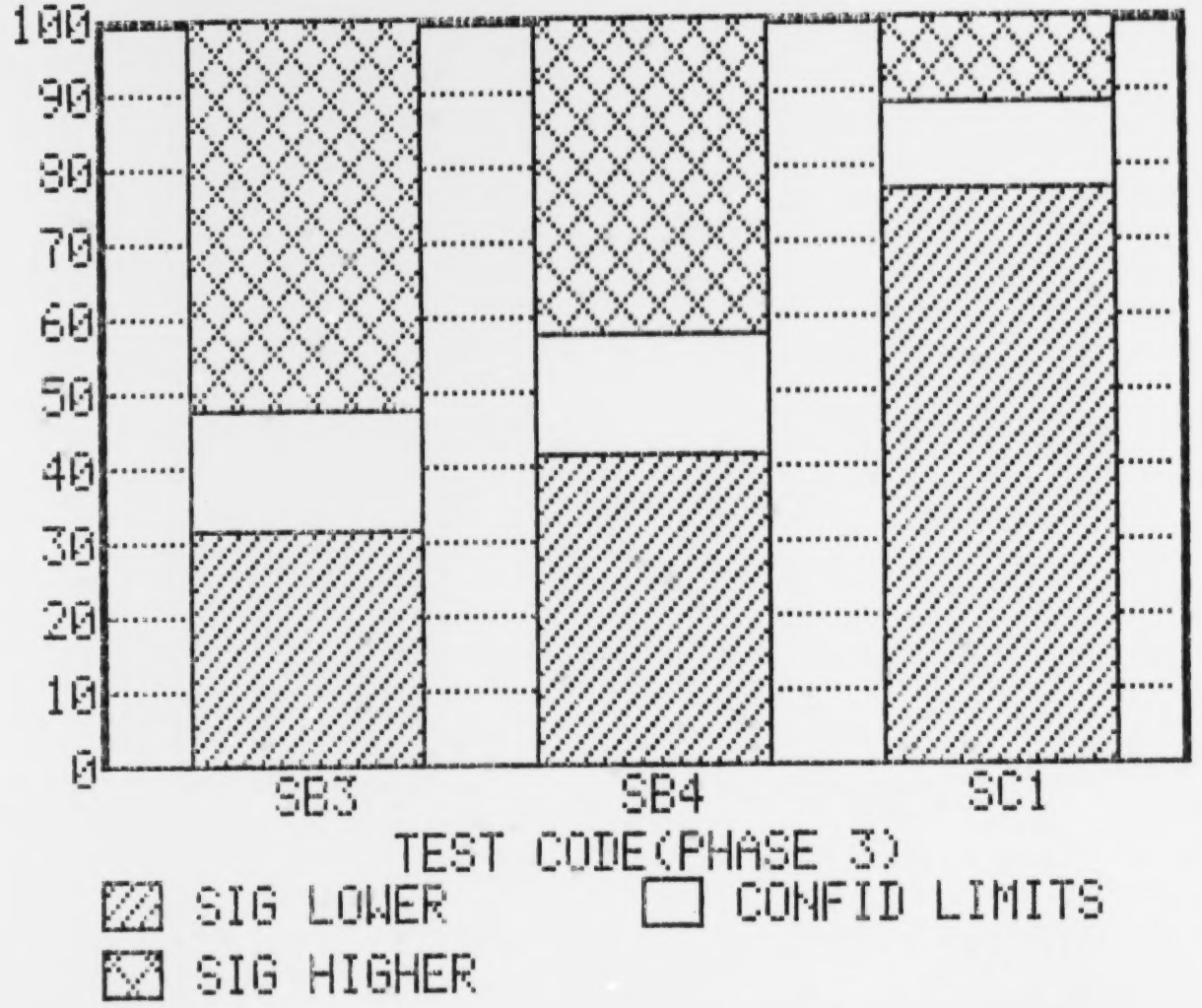

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
SB3
SB4
SC1
TEST CODE(PHASE 3)
SIG LOWER
CONFID LIMITS
SIG HIGHER

FIG 1.2 MEASUREMENT

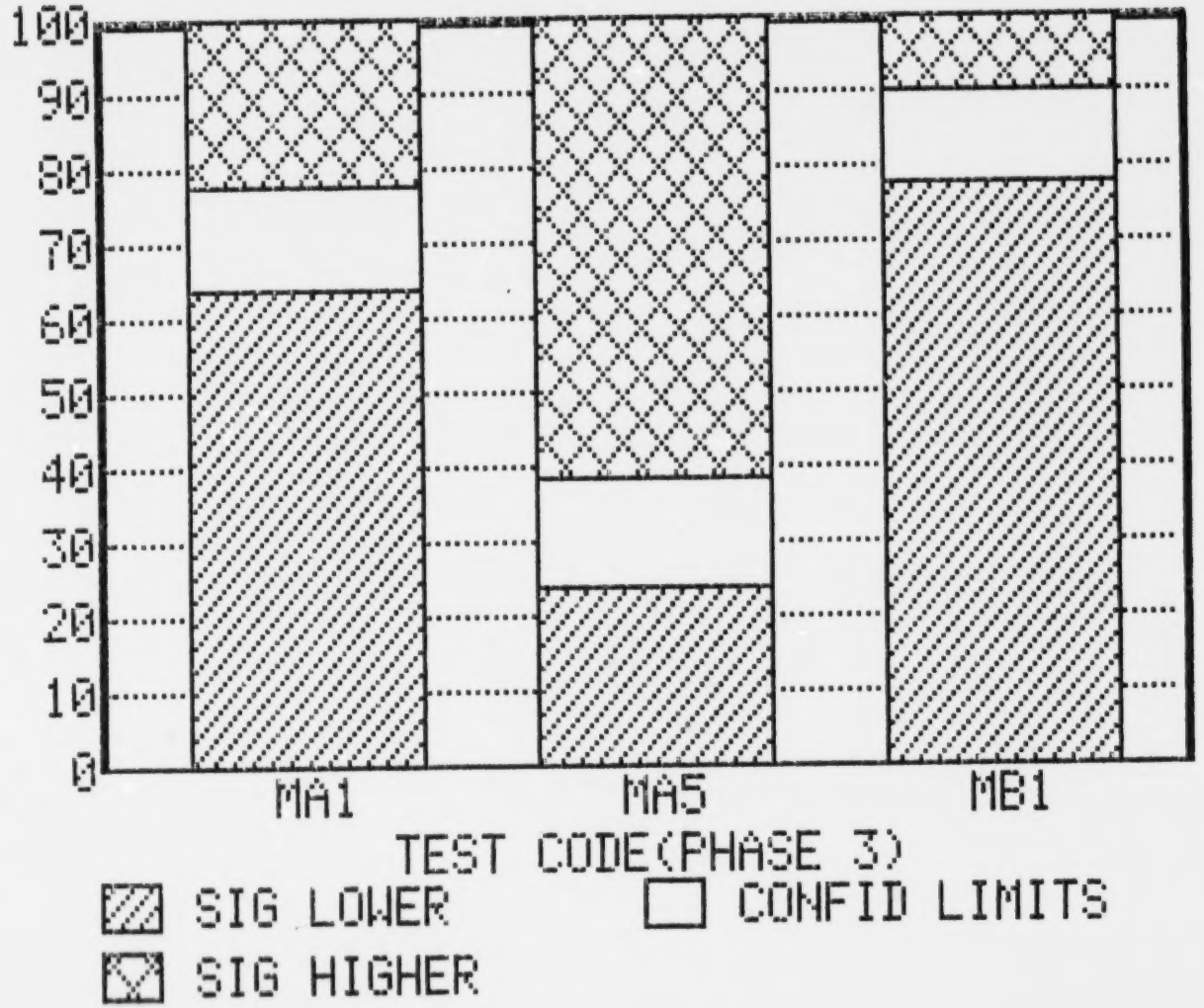


FIG 1.3 MEASUREMENT

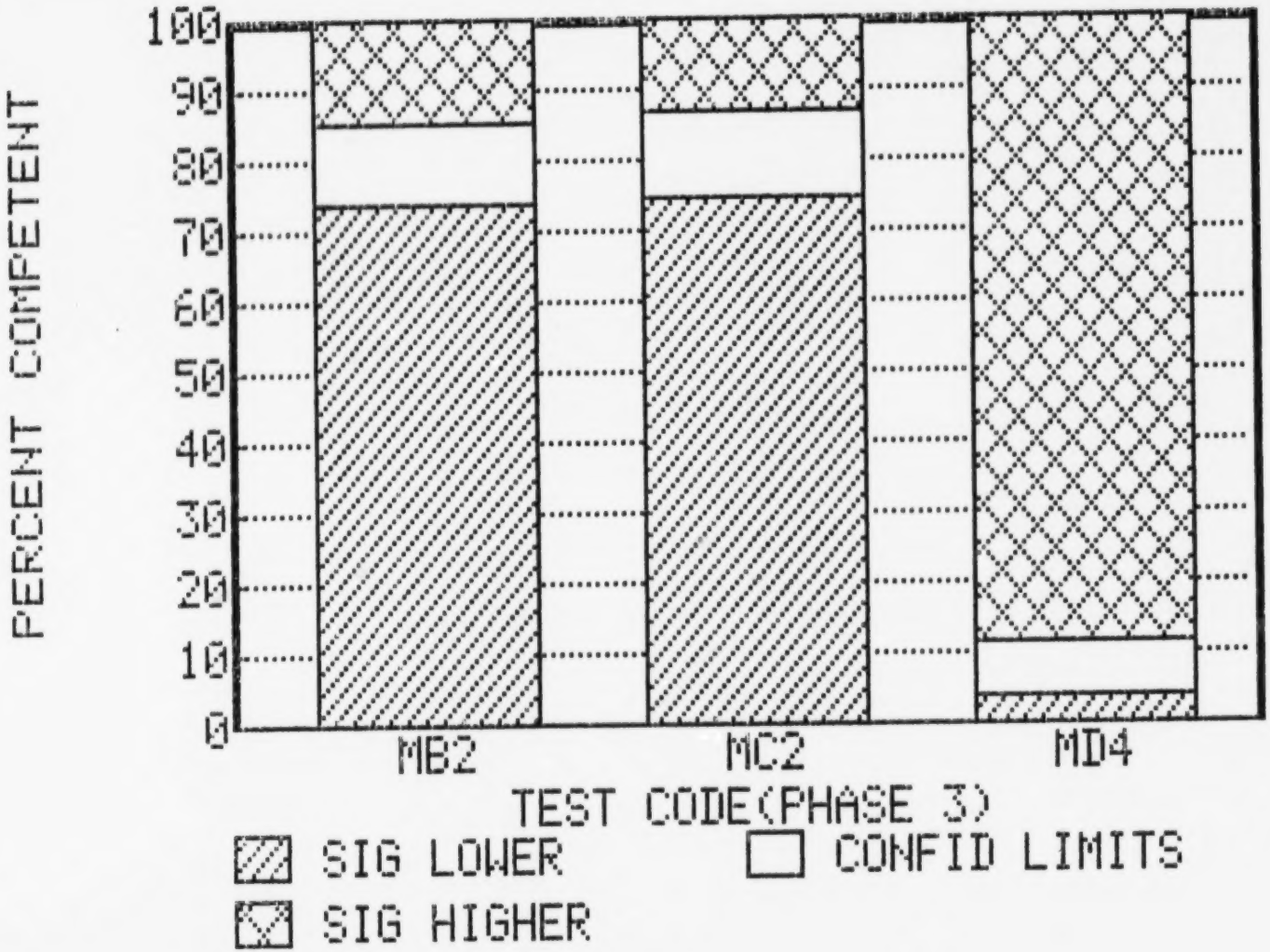


FIG 1.4 MEASUREMENT

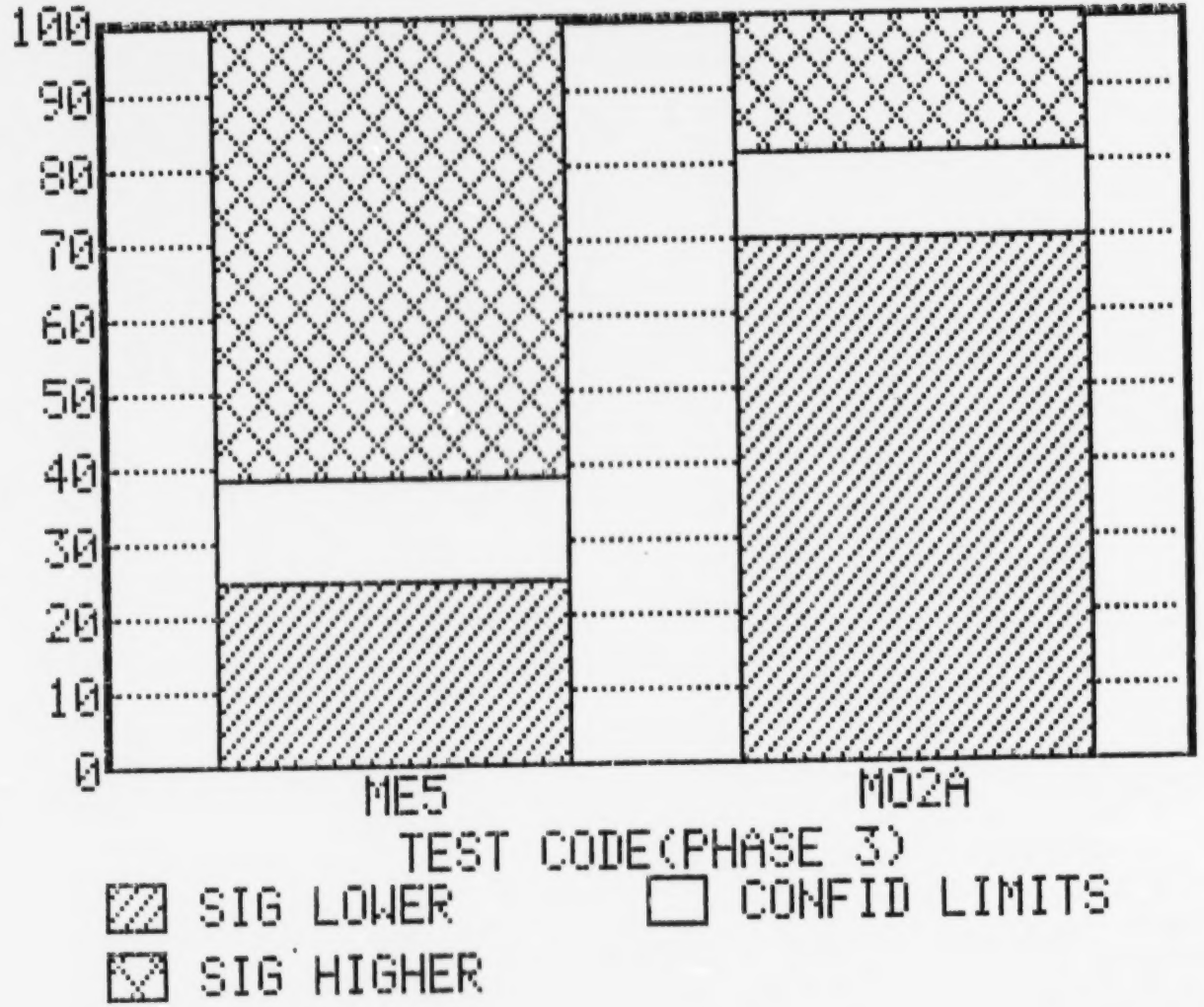


FIG 1.5 MEASUREMENT

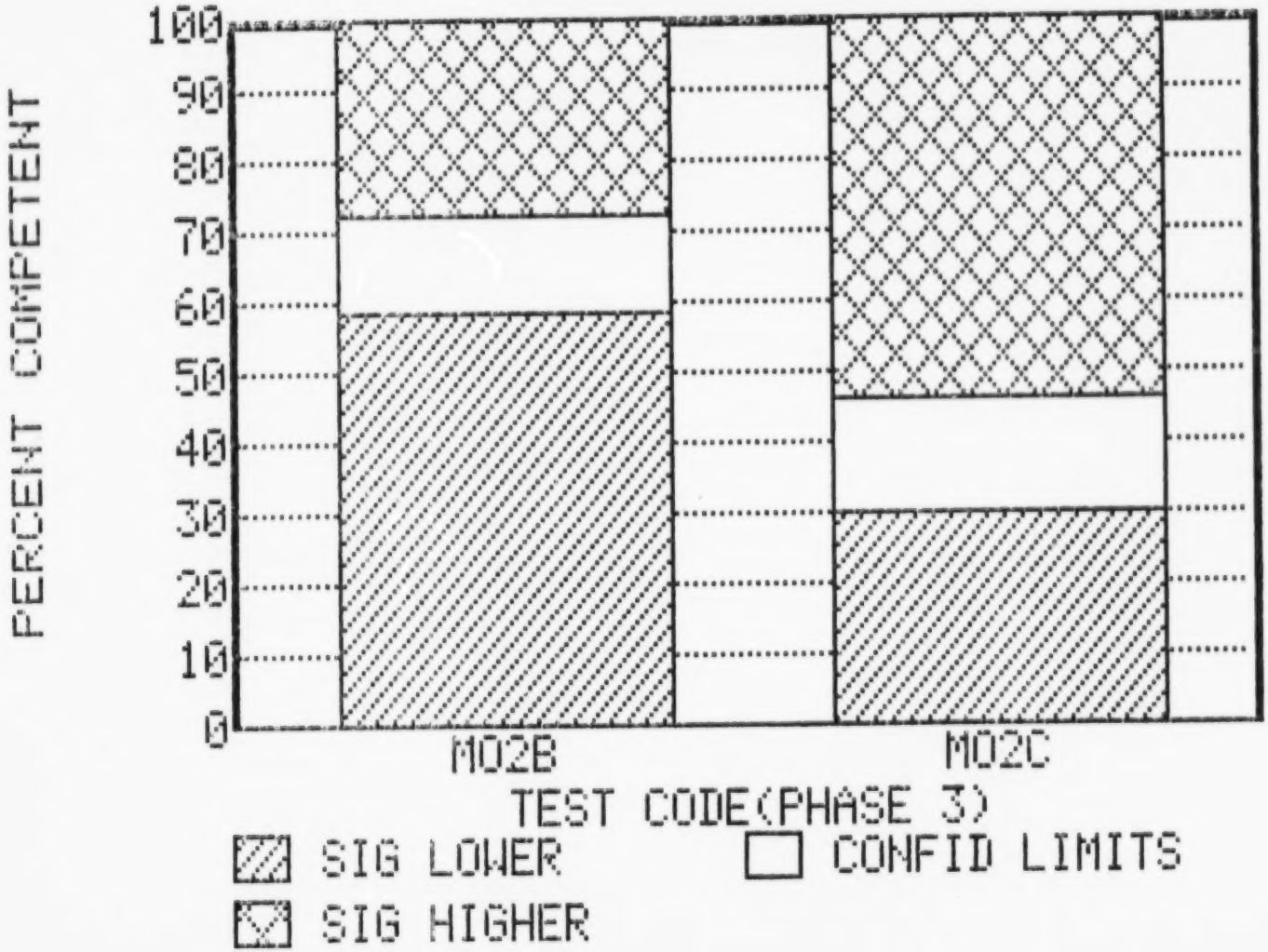


FIG 1.6 ADD AND SUBTRACT

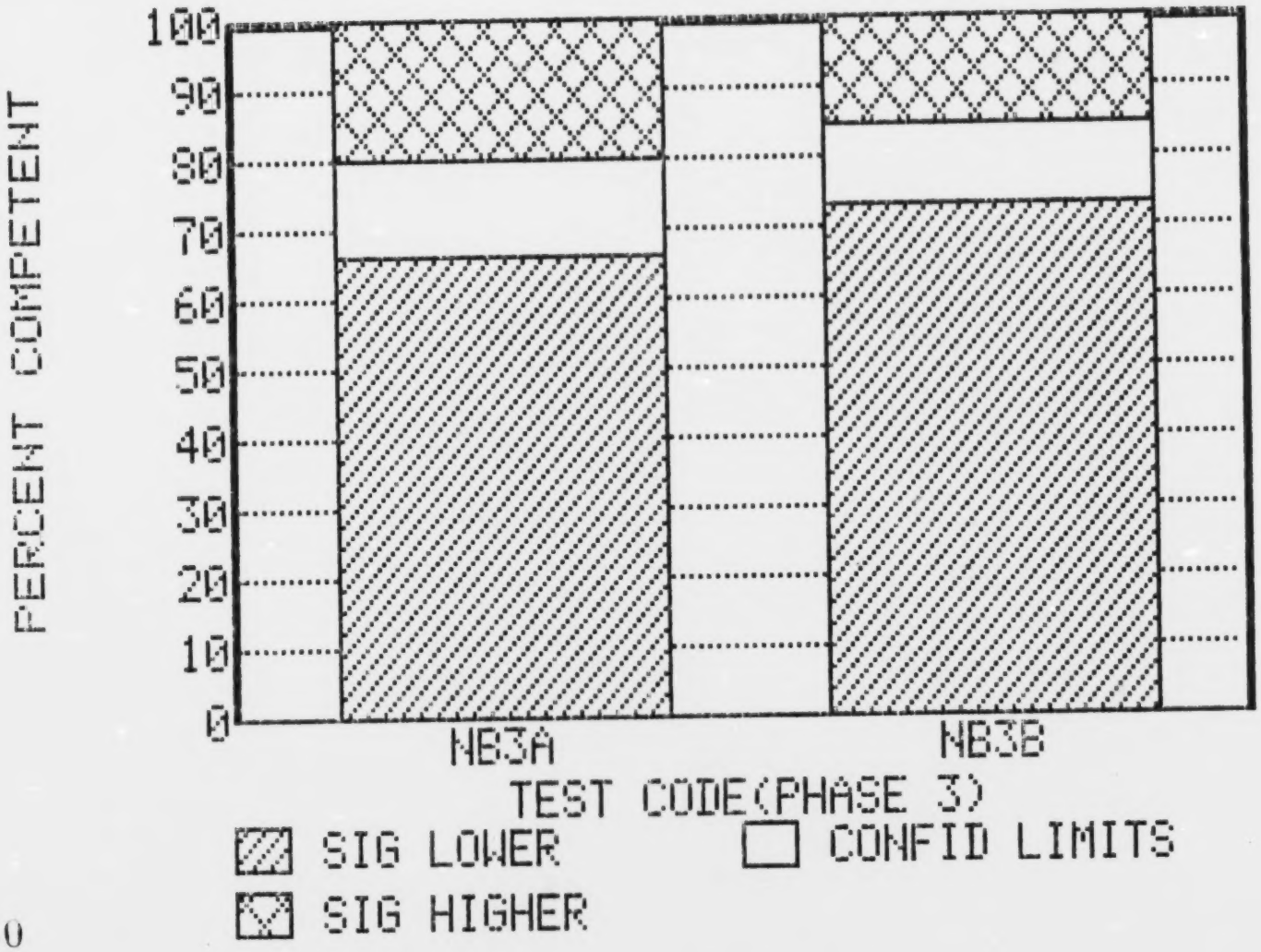


FIG 1.7 ADD AND SUBTRACT

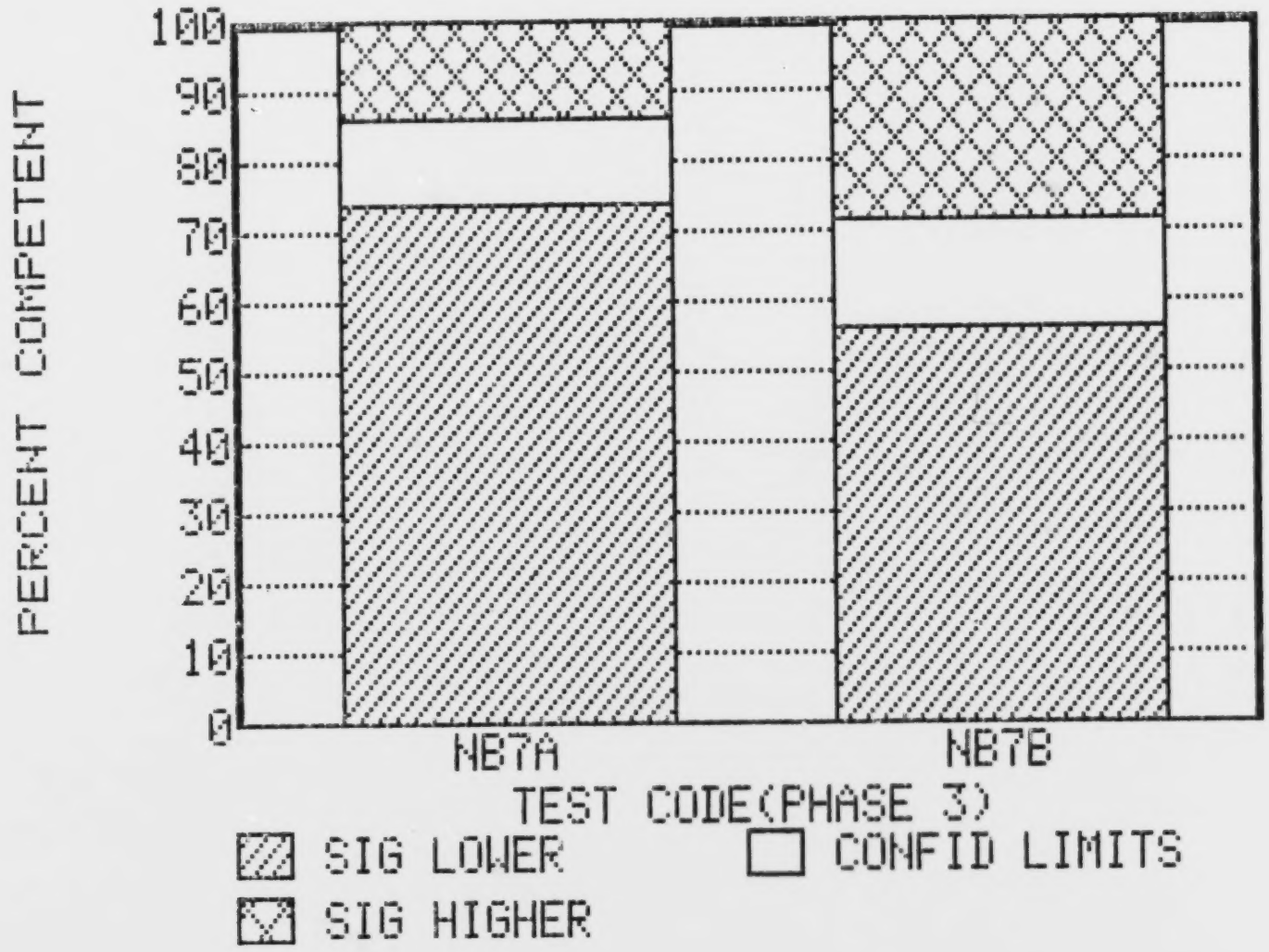


FIG 1.8 MULTIPLY & DIVIDE

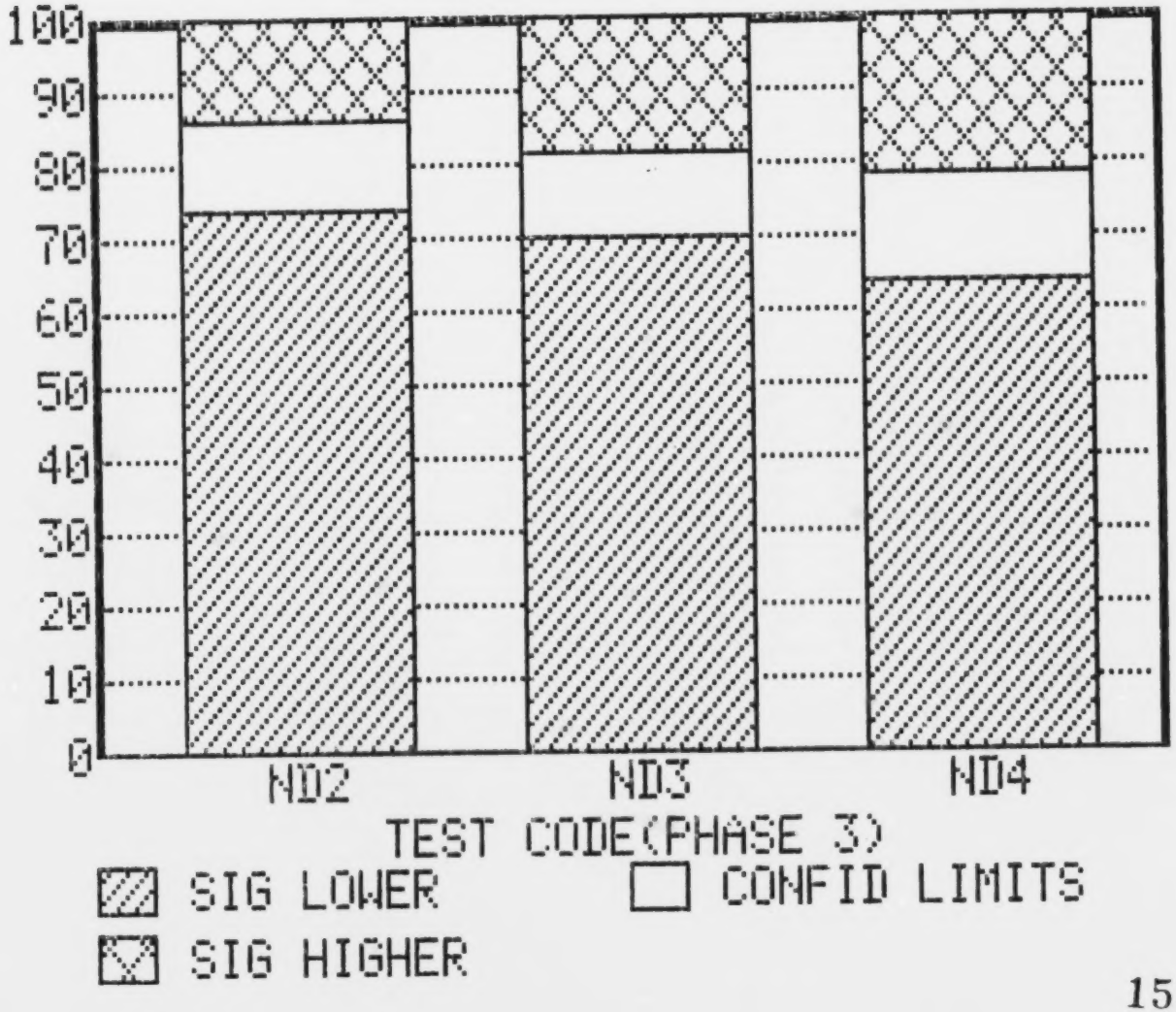


FIG 1.9 MULTIPLY & DIVIDE

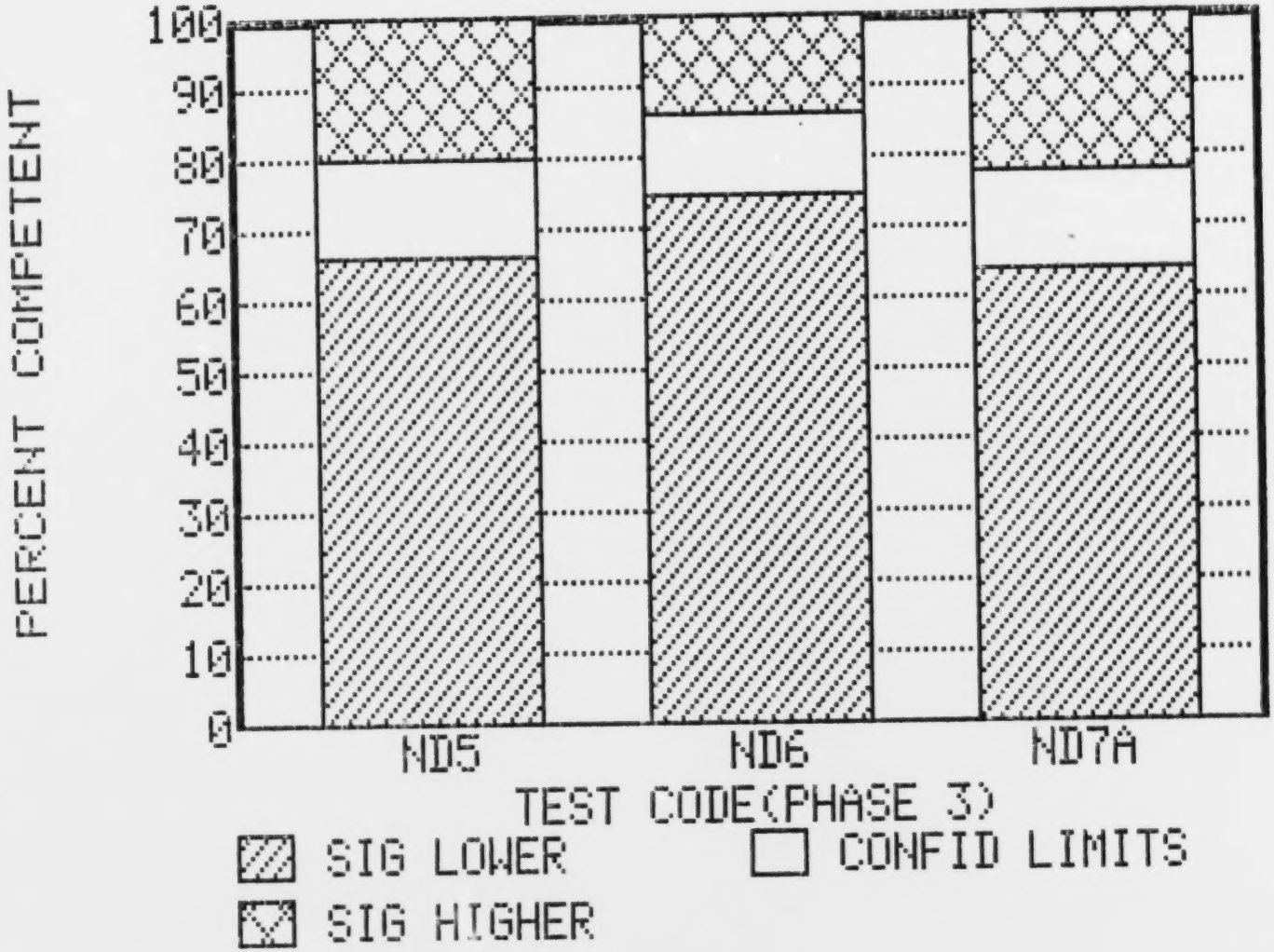


FIG 1.10 MULTIPLY&DIVIDE

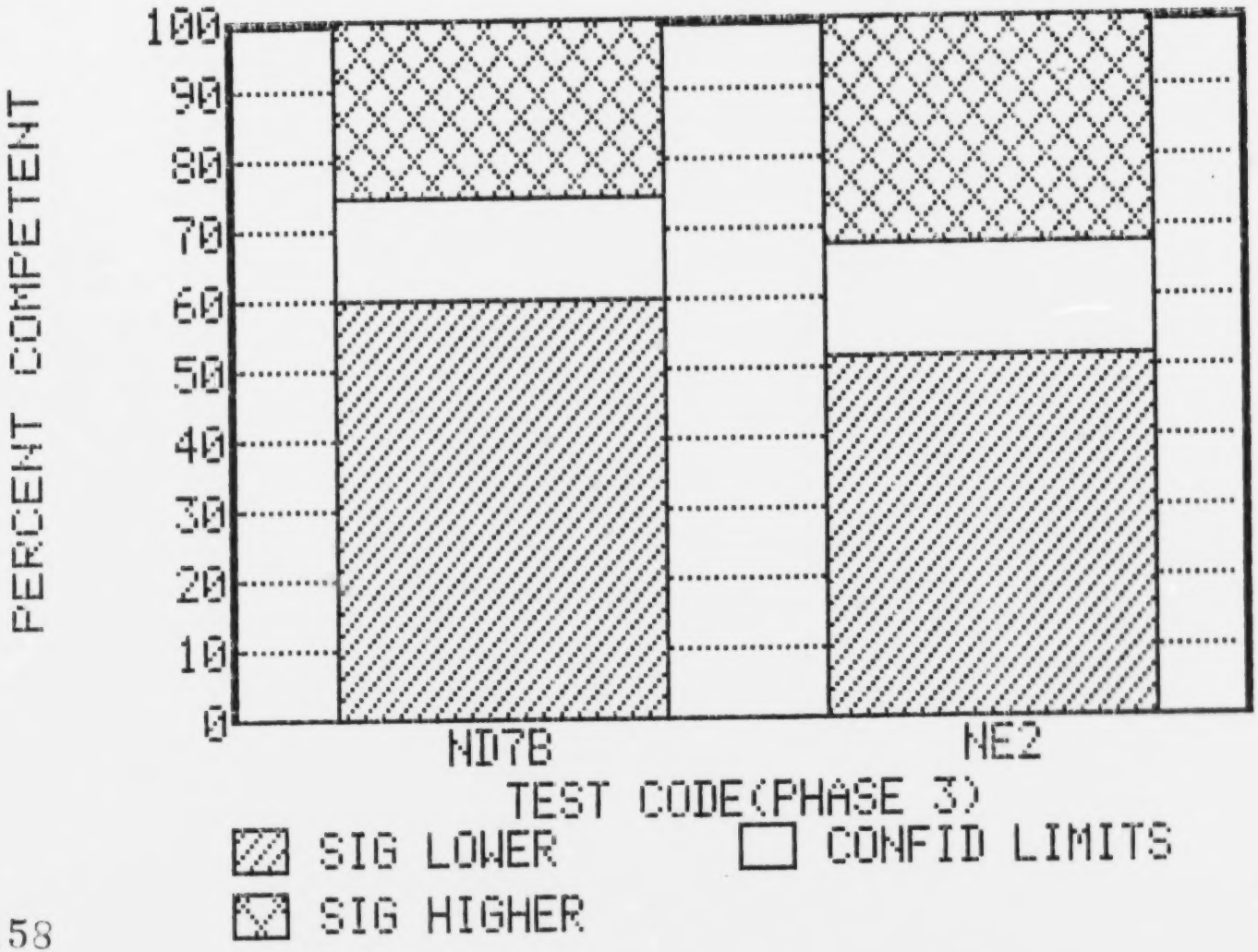


FIG 1.11 NUMBER RELATIONS

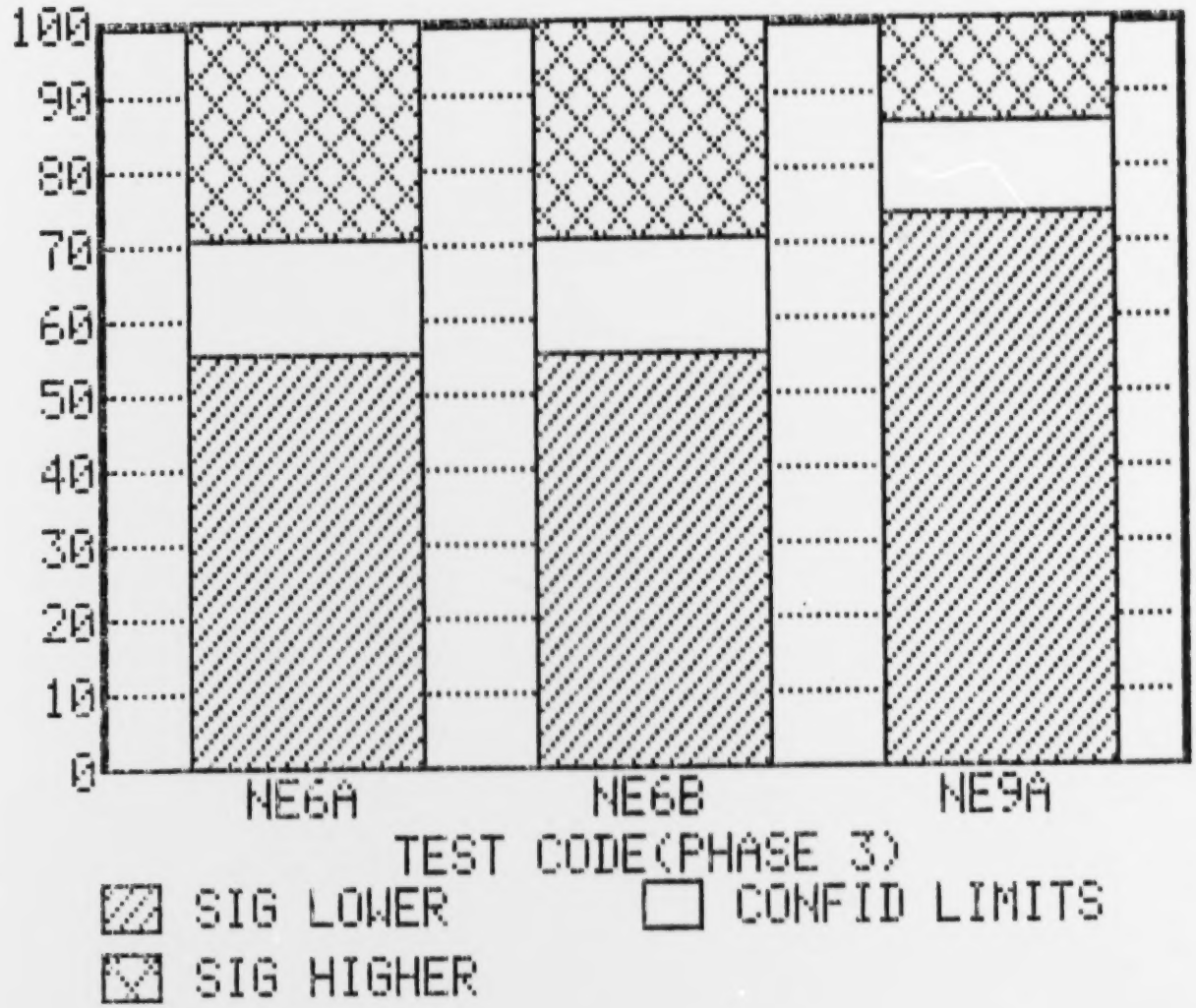


FIG 2.1 SPACE/MEARSUREMENT

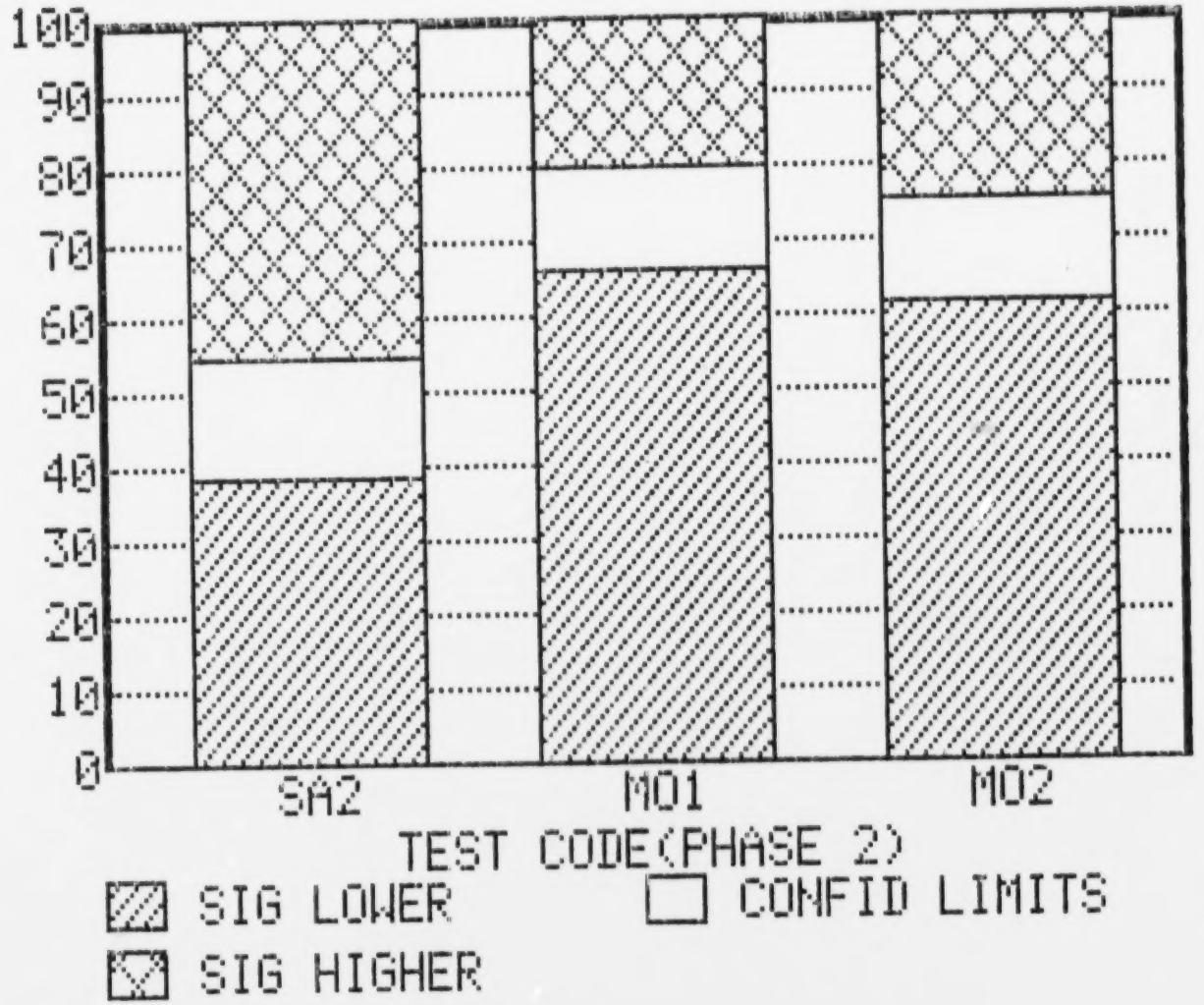


FIG 2.2 MEASUREMENT

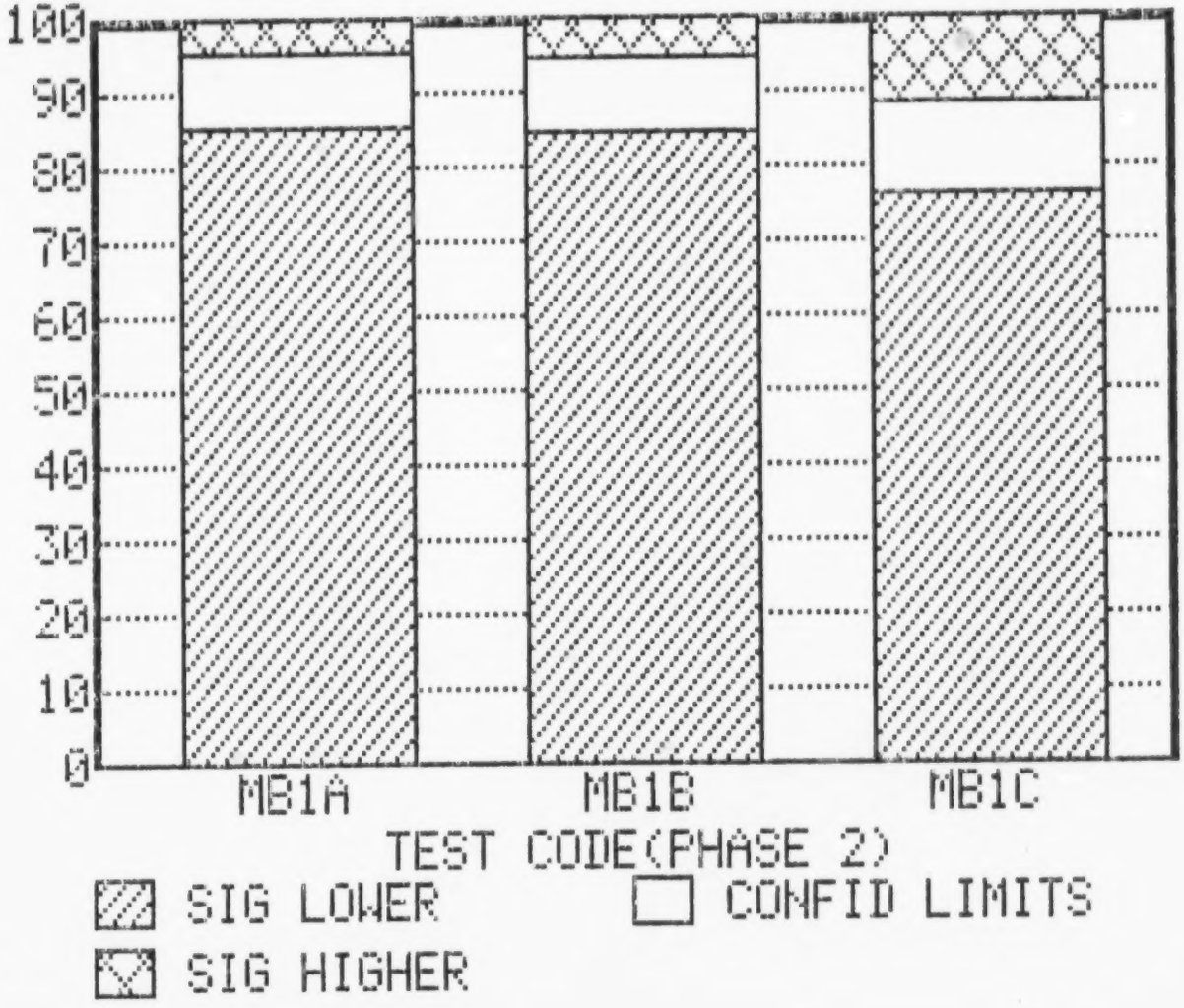


FIG 2.3 MEASUREMENT

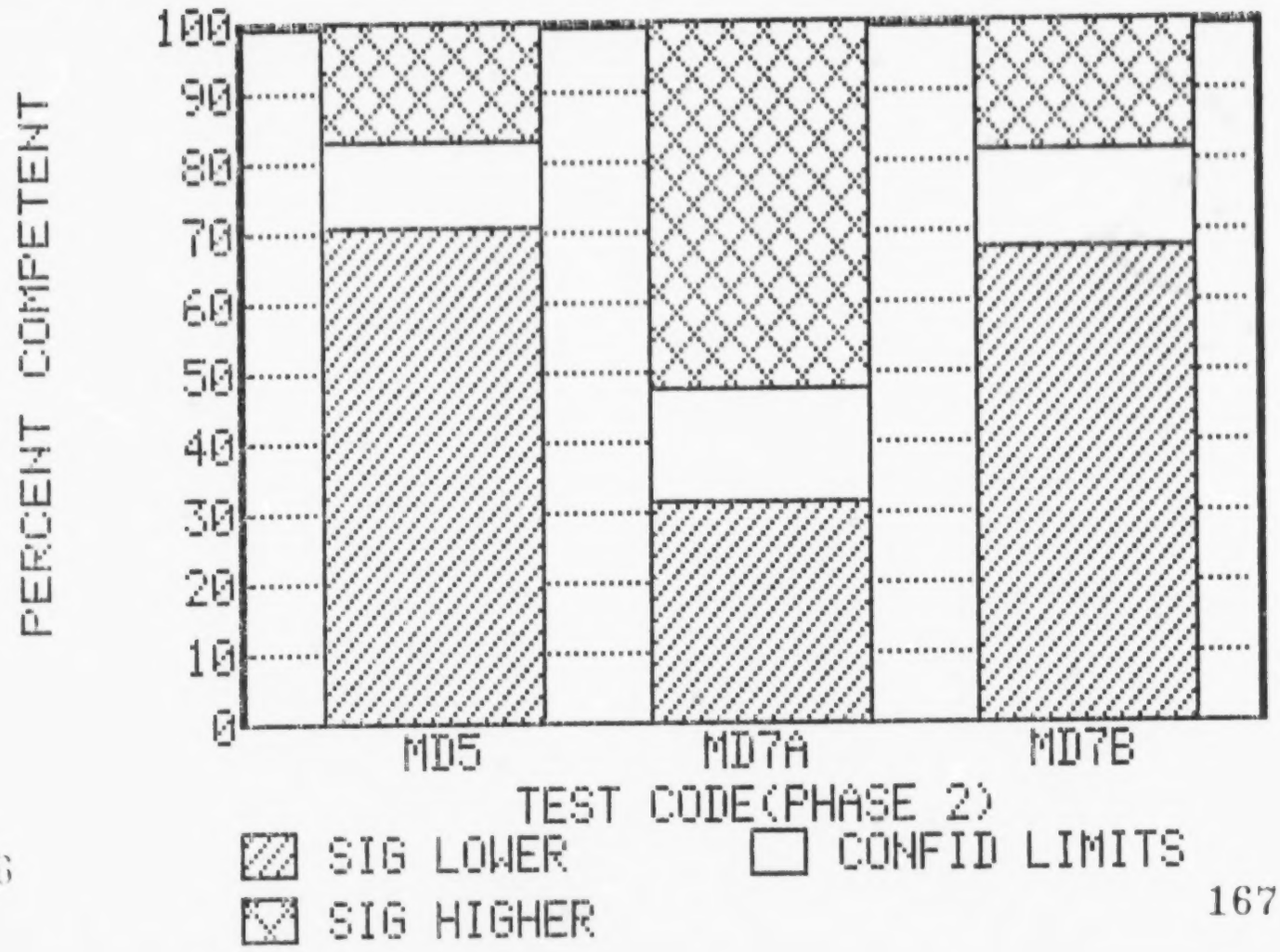


FIG 2.4 ADD AND SUBTRACT

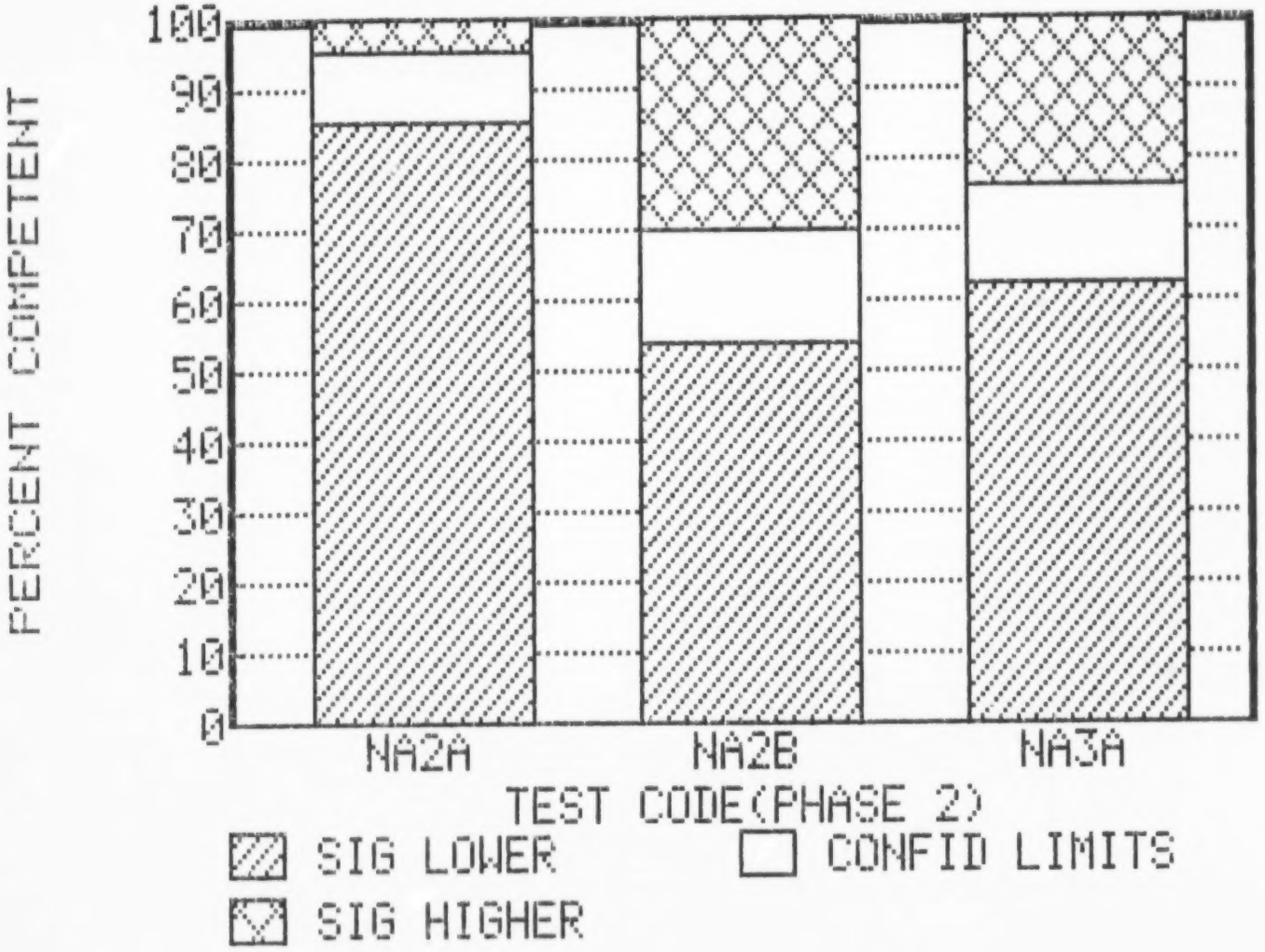


FIG 2.5 ADD AND SUBTRACT

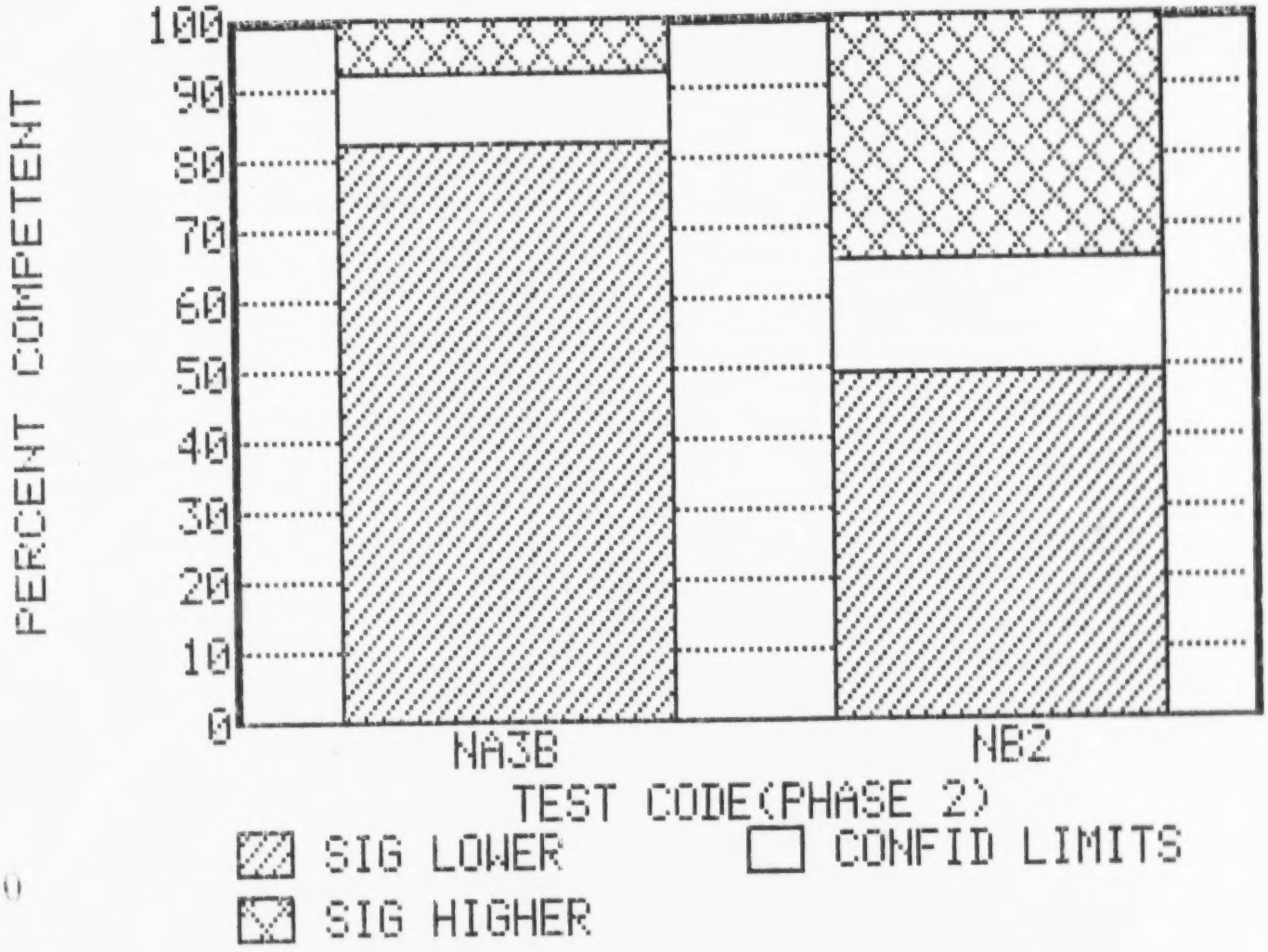


FIG 2.6 ADD AND SUBTRACT

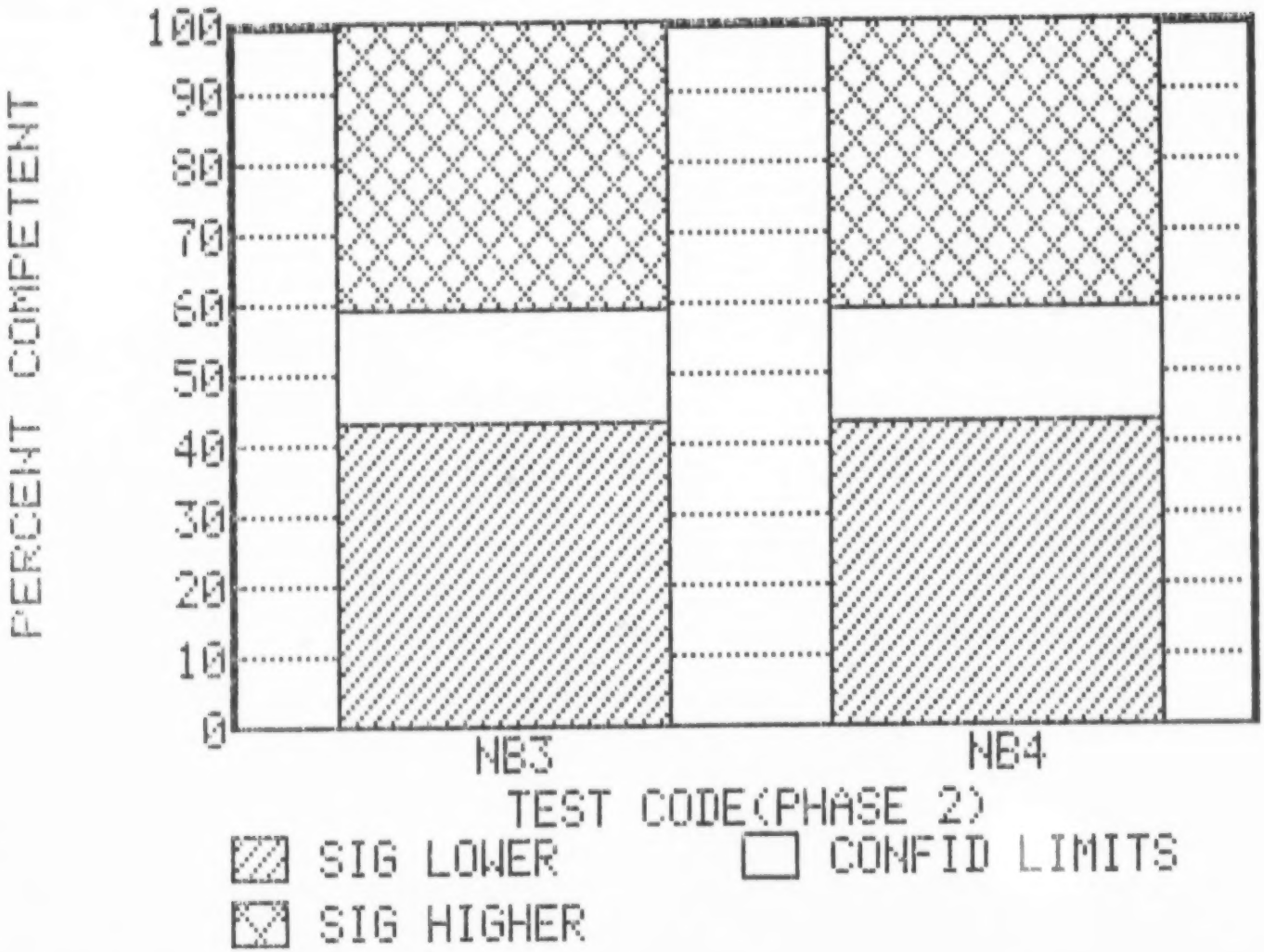


FIG 2.7 MULTIPLY & DIVIDE

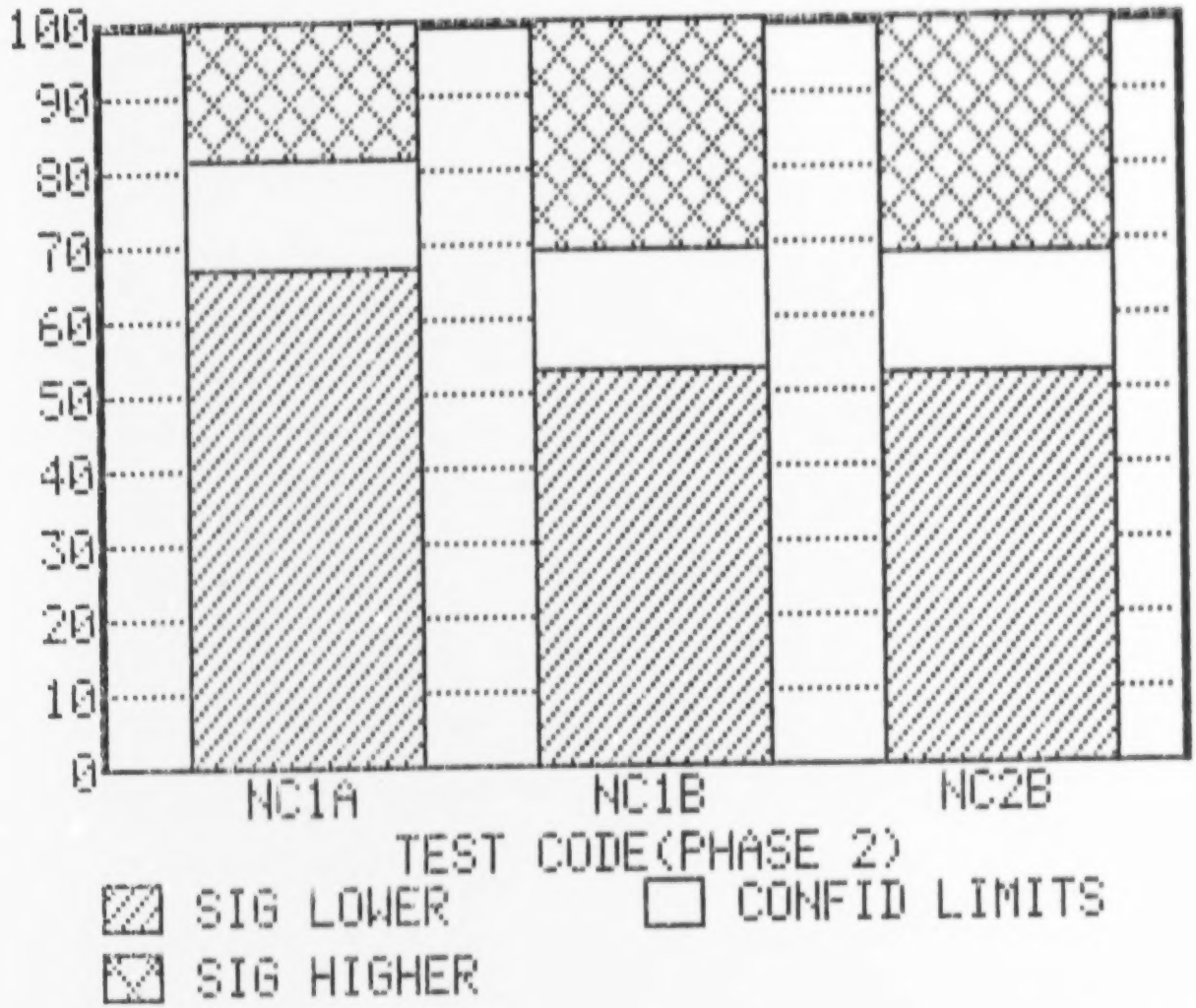
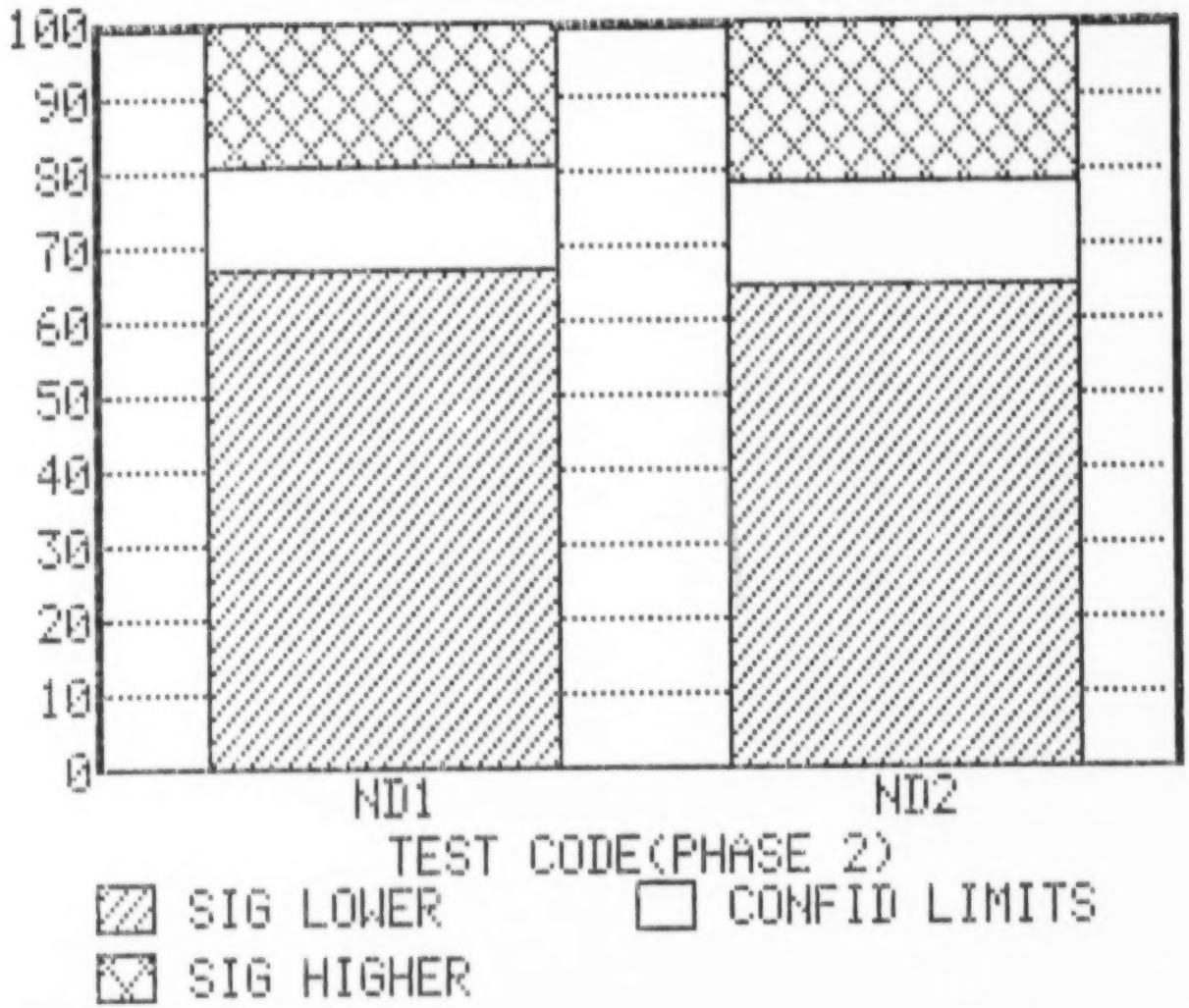


FIG 2.8 MULTIPLY & DIVIDE



176

177

FIG 2.9 MULTIPLY & DIVIDE

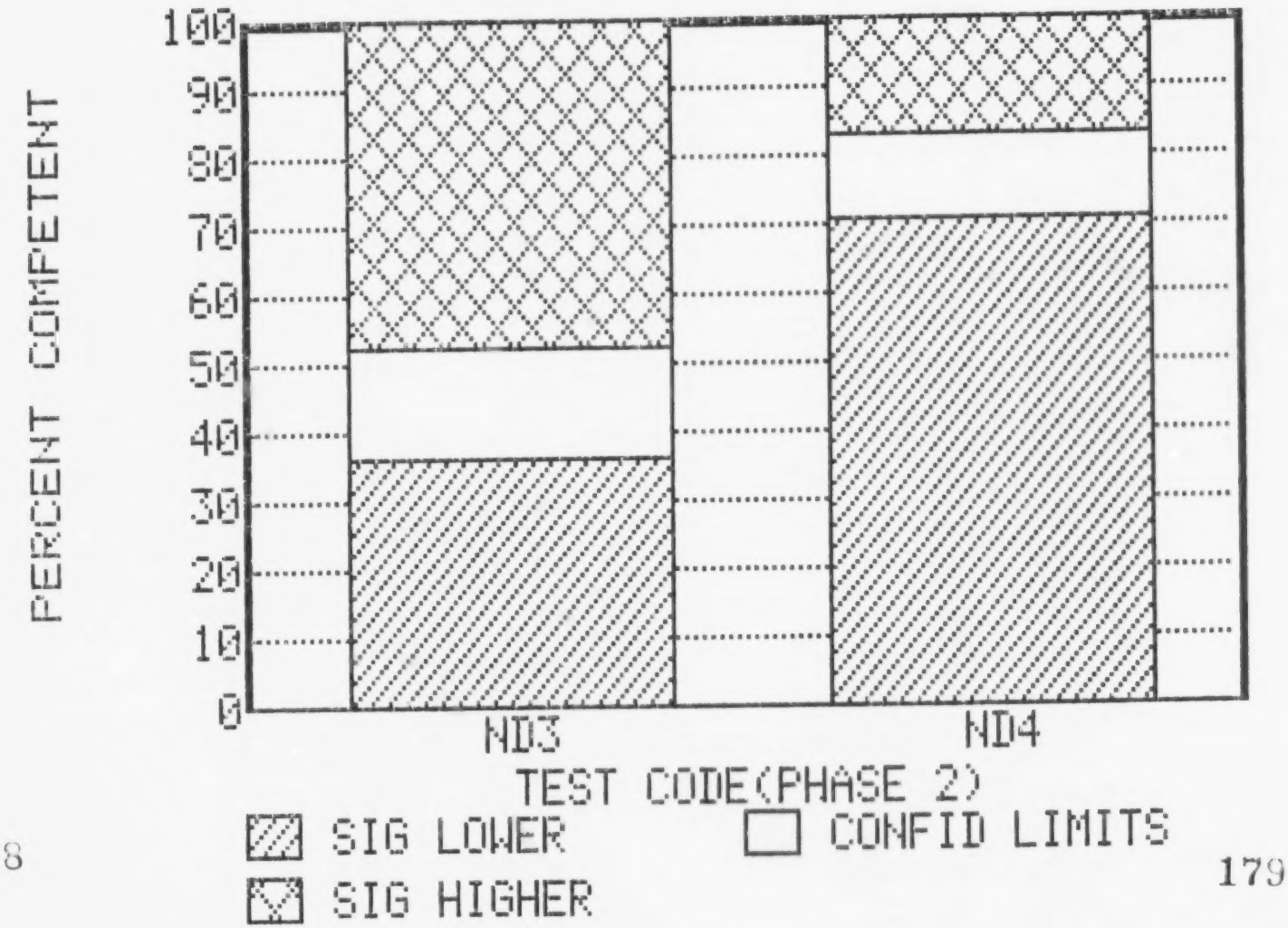


FIG 2.10 RELATIONS

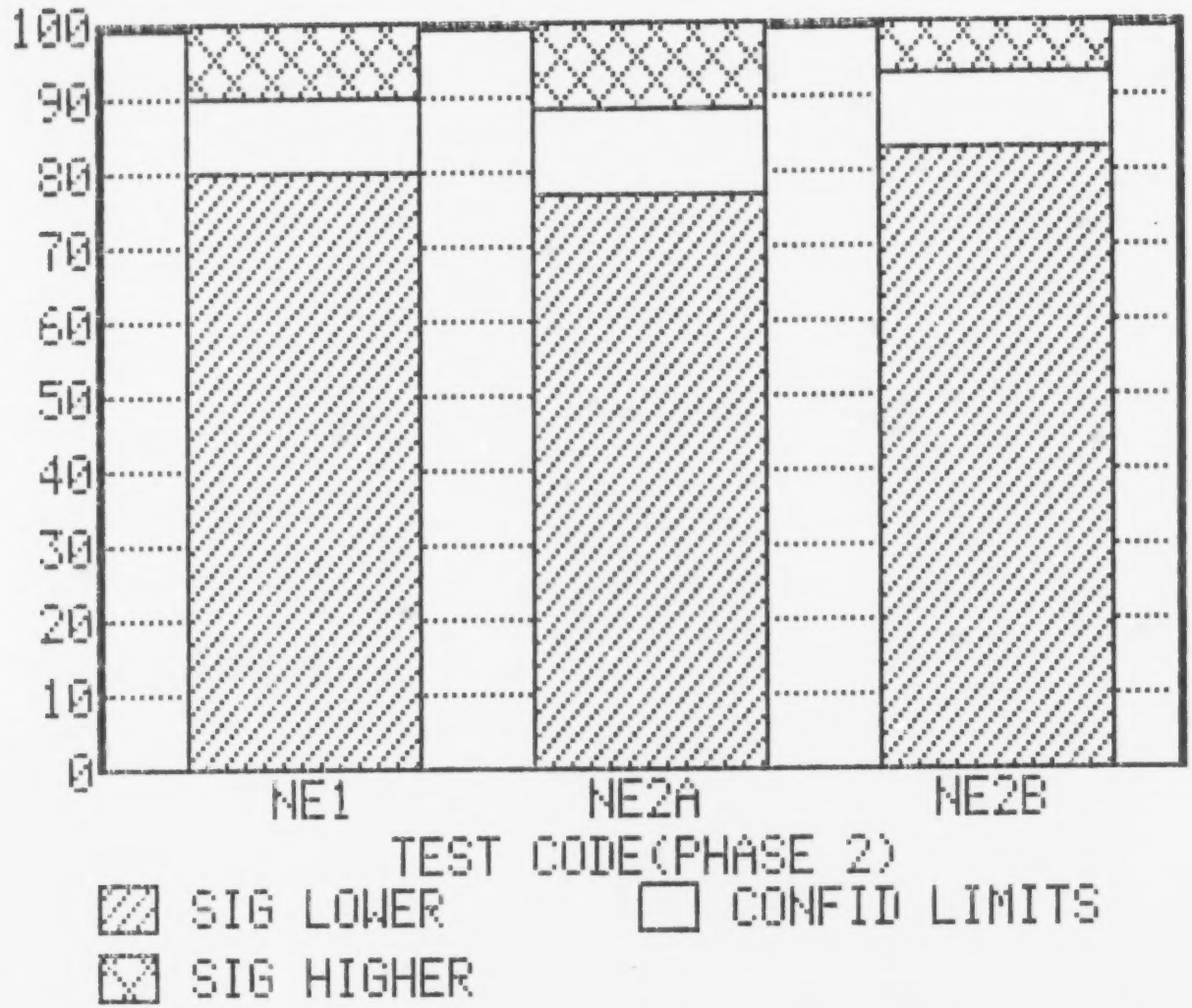


FIG 2.11 RELATIONS

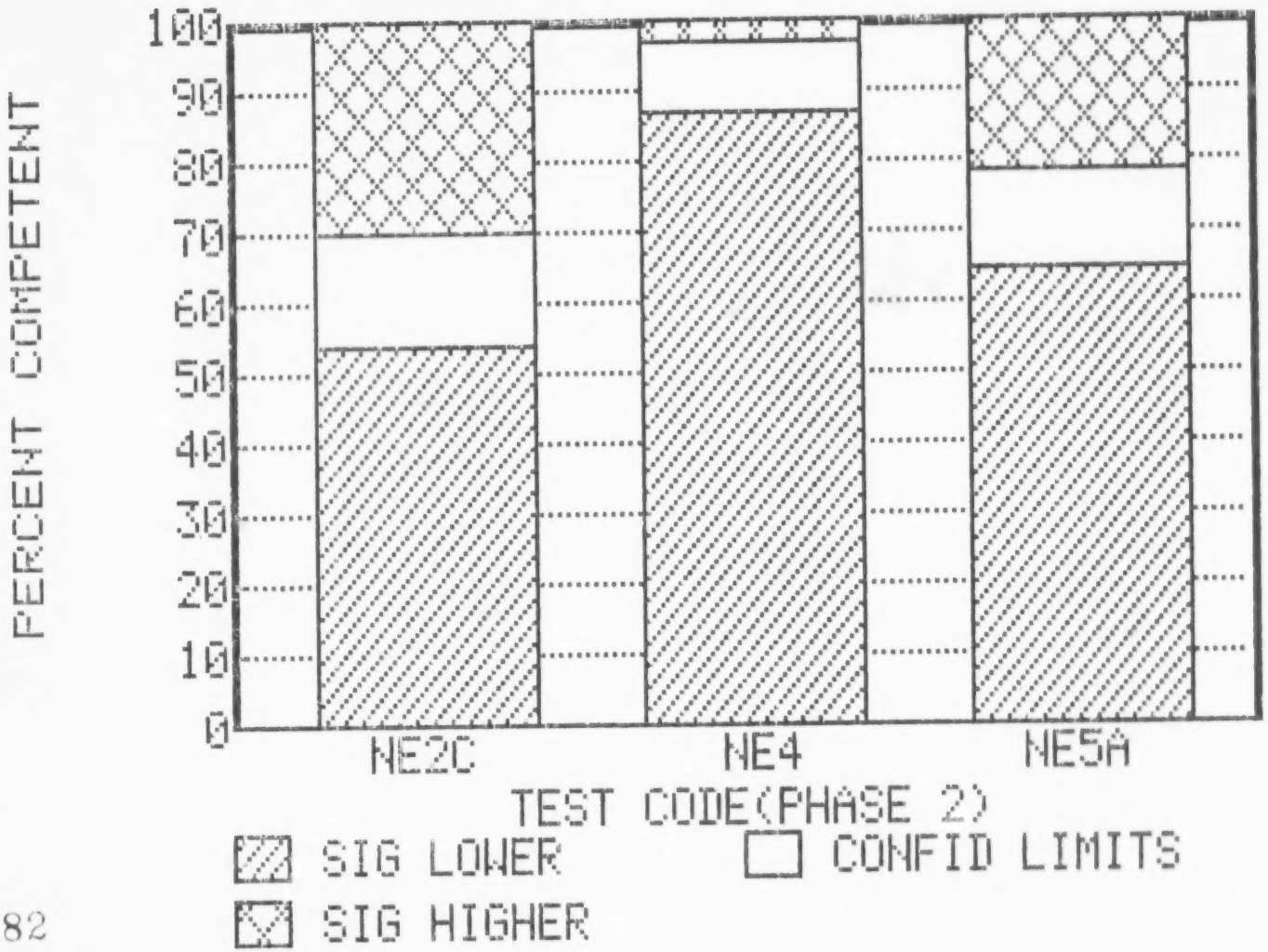


FIG 2.12 RELATIONS

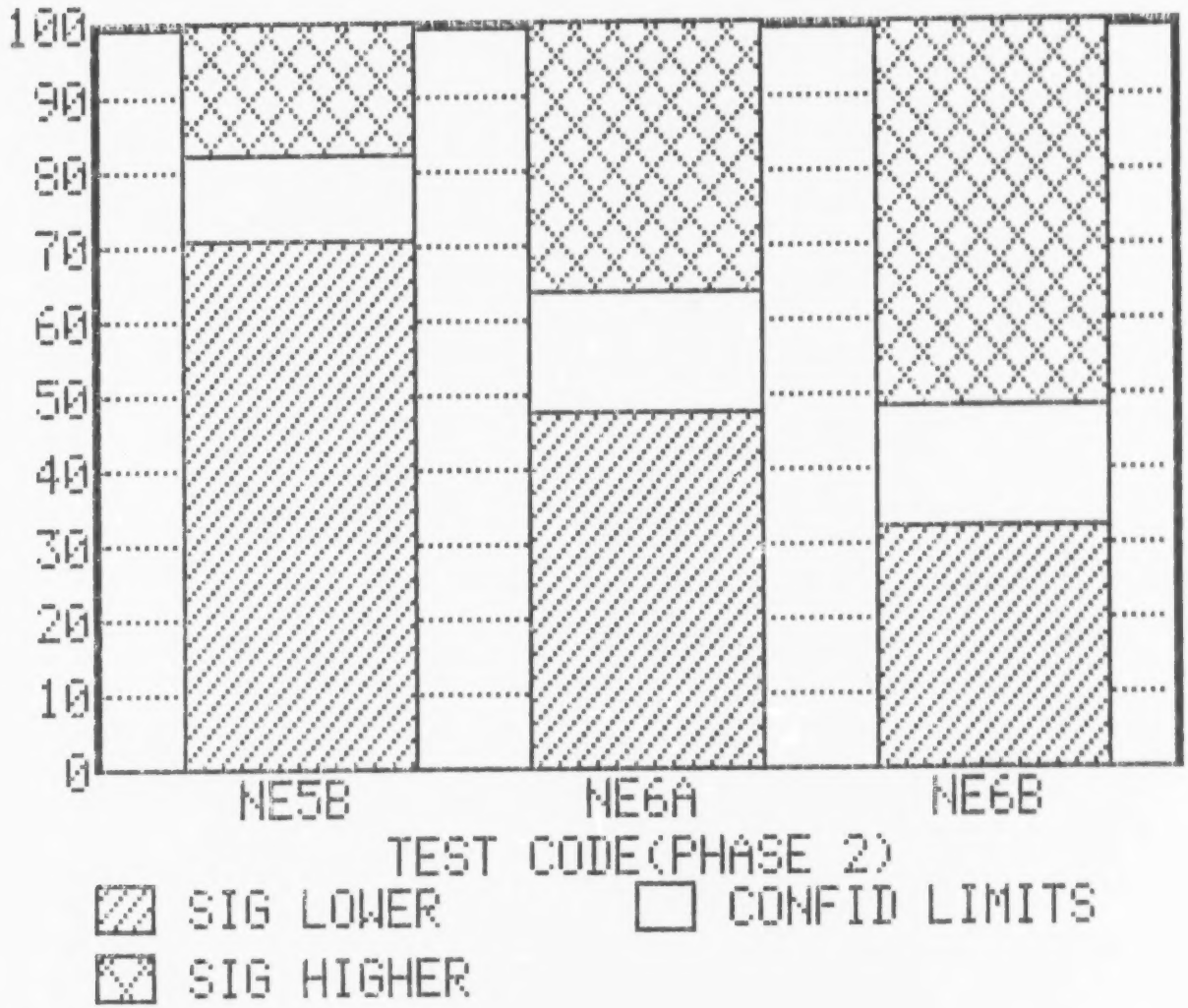


FIG 2.13 RELATIONS

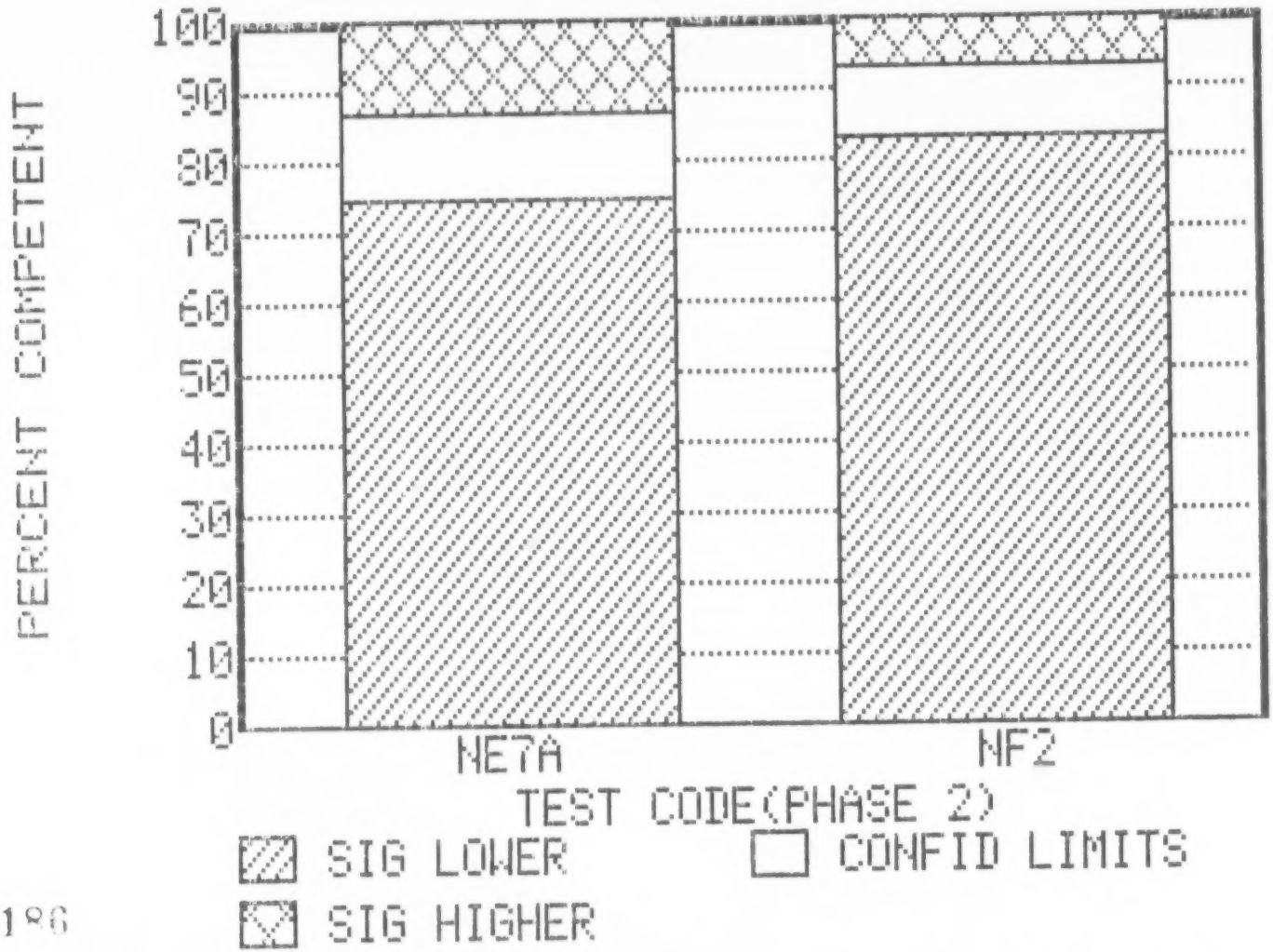


FIG 3.1 COMPREHENSION

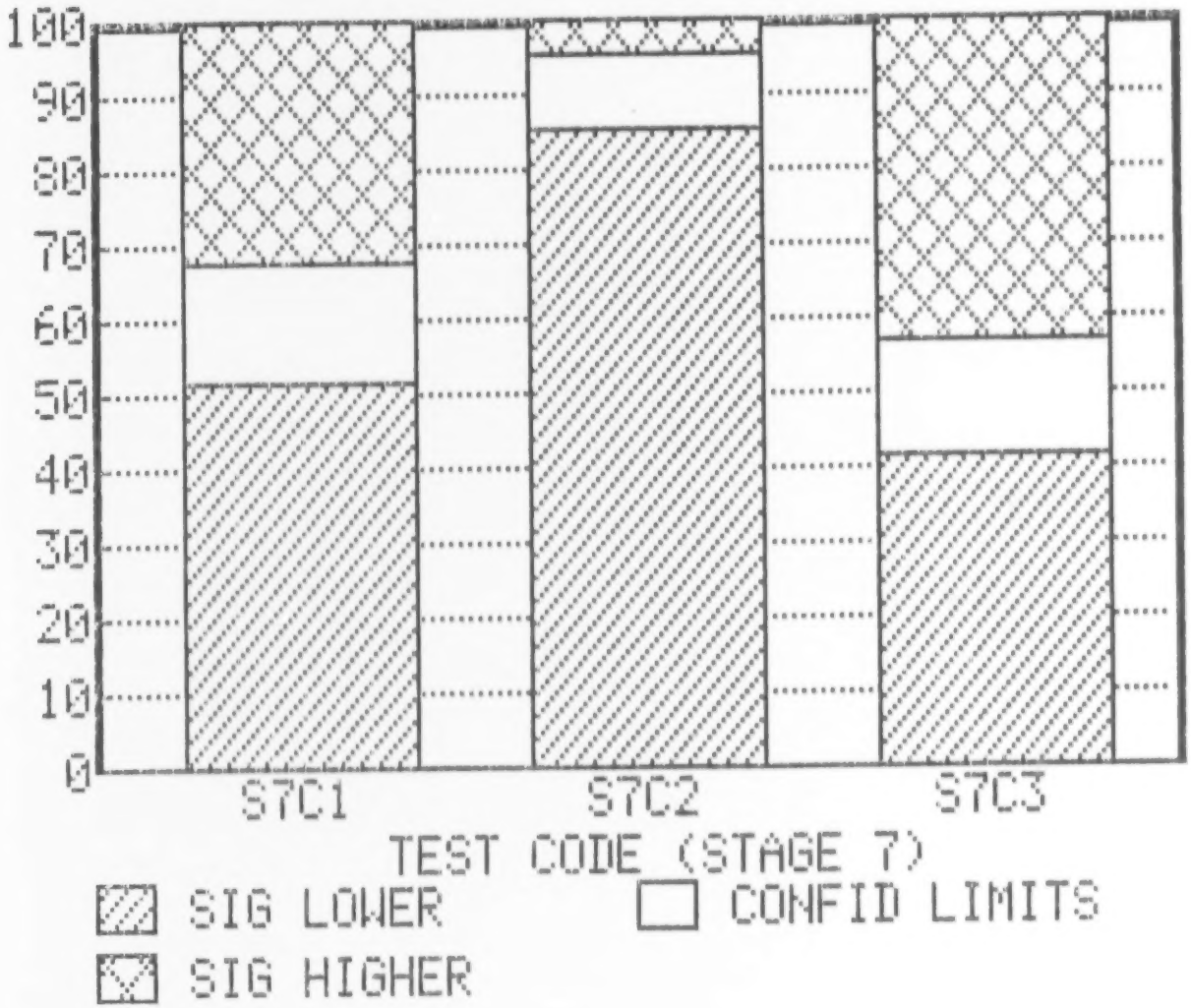


FIG 3.2 COMPREHENSION

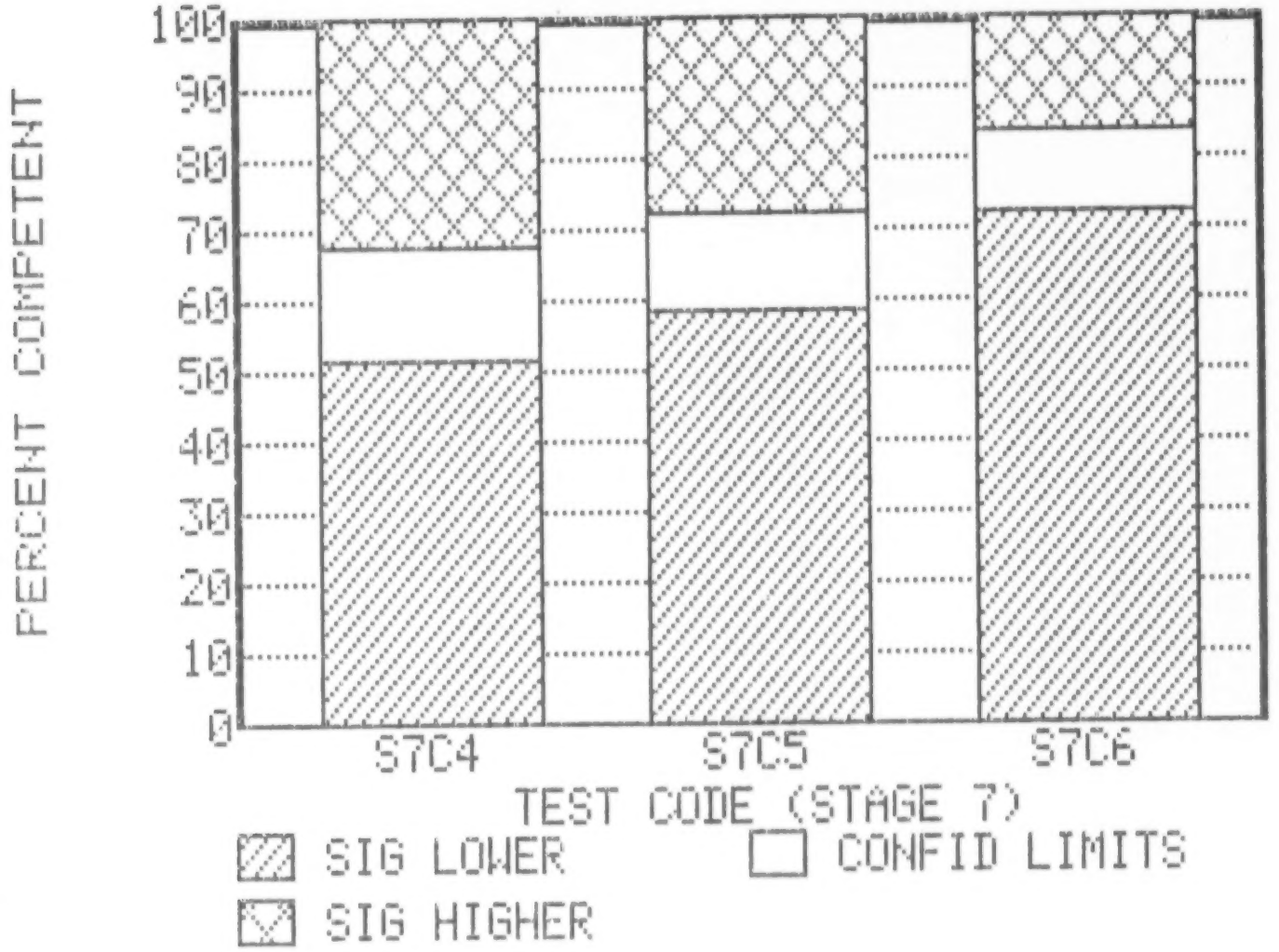


FIG 3.3 COMPREHENSION

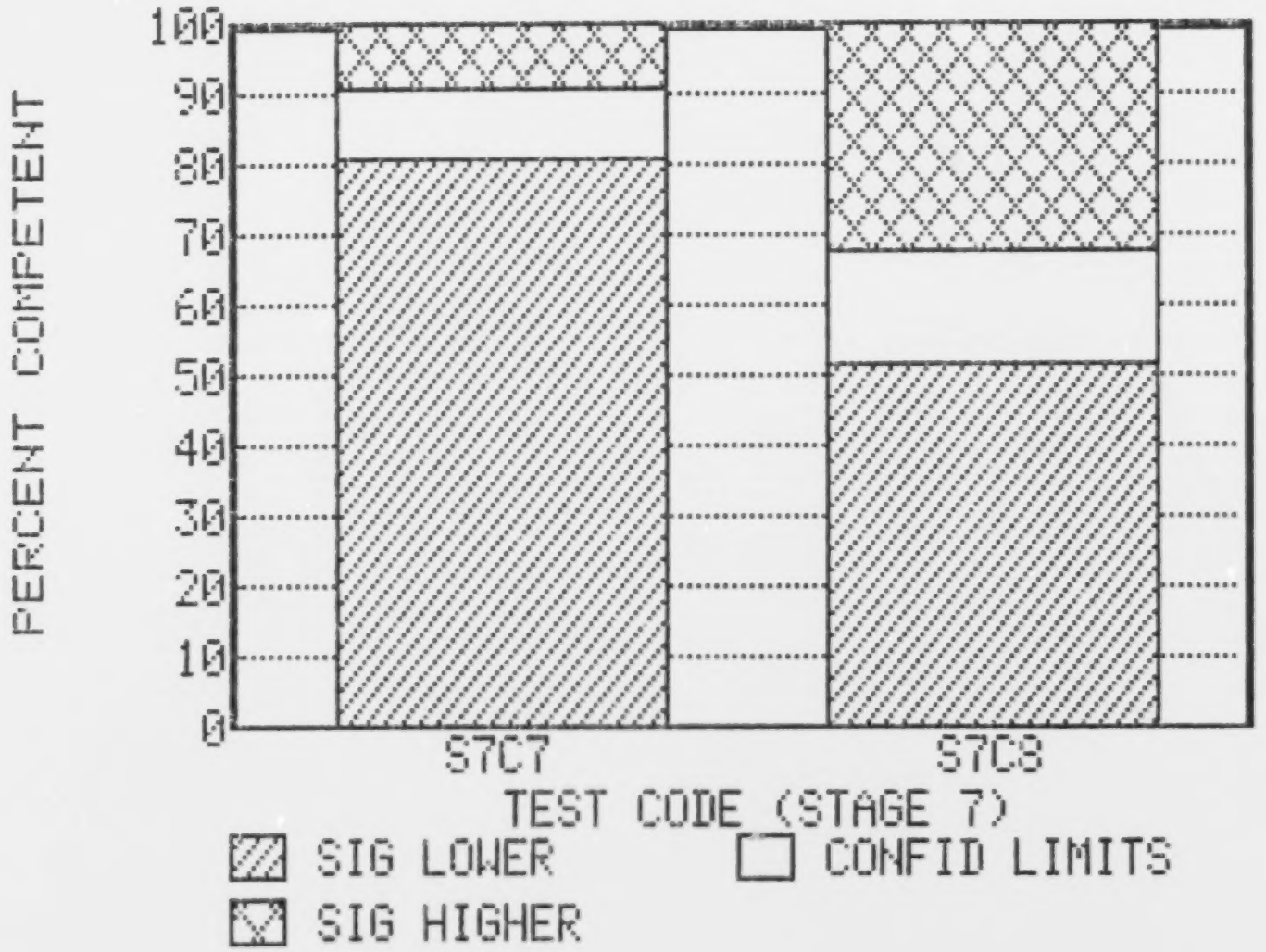


FIG 3.4 DICTIONARY SKILLS

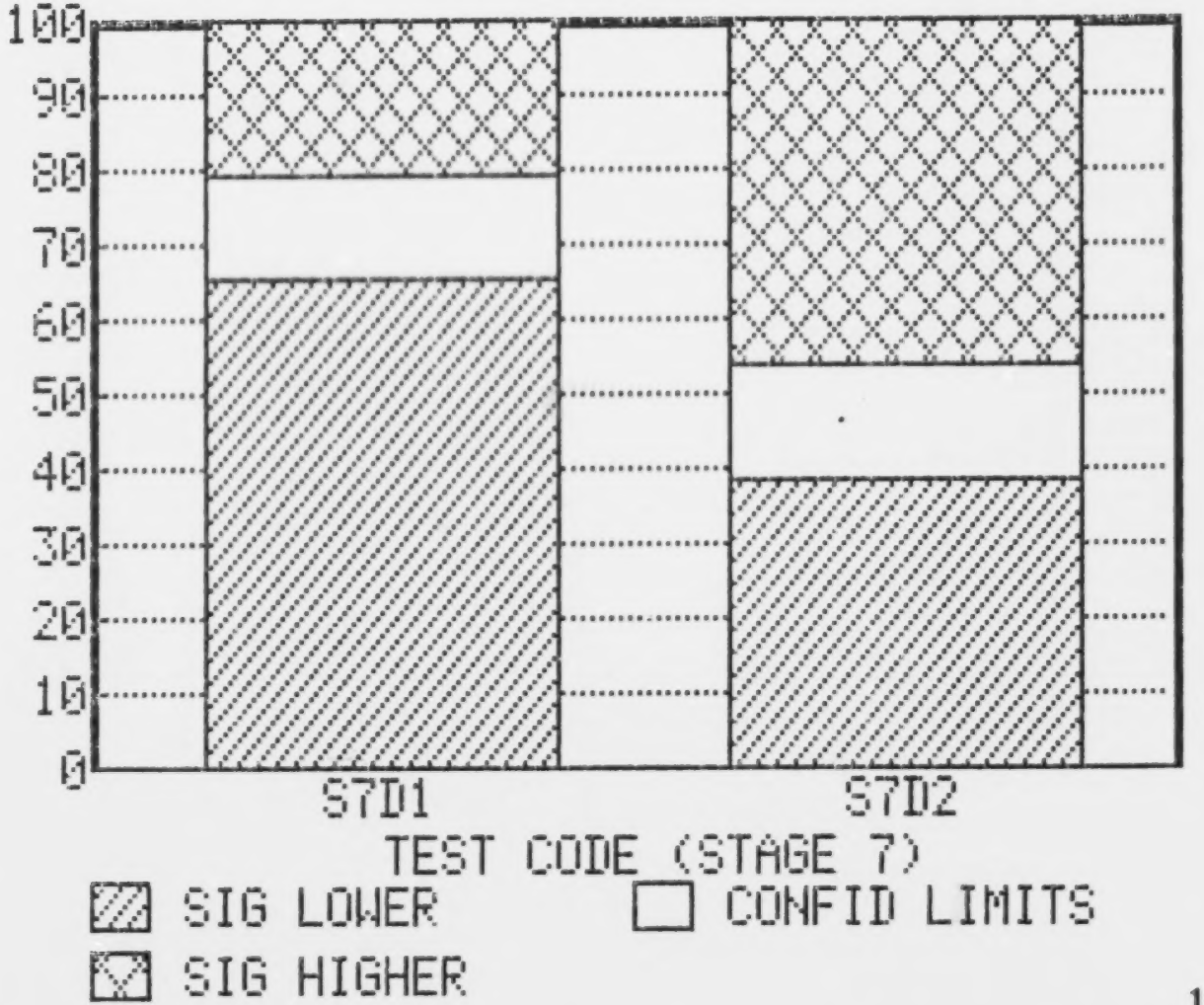
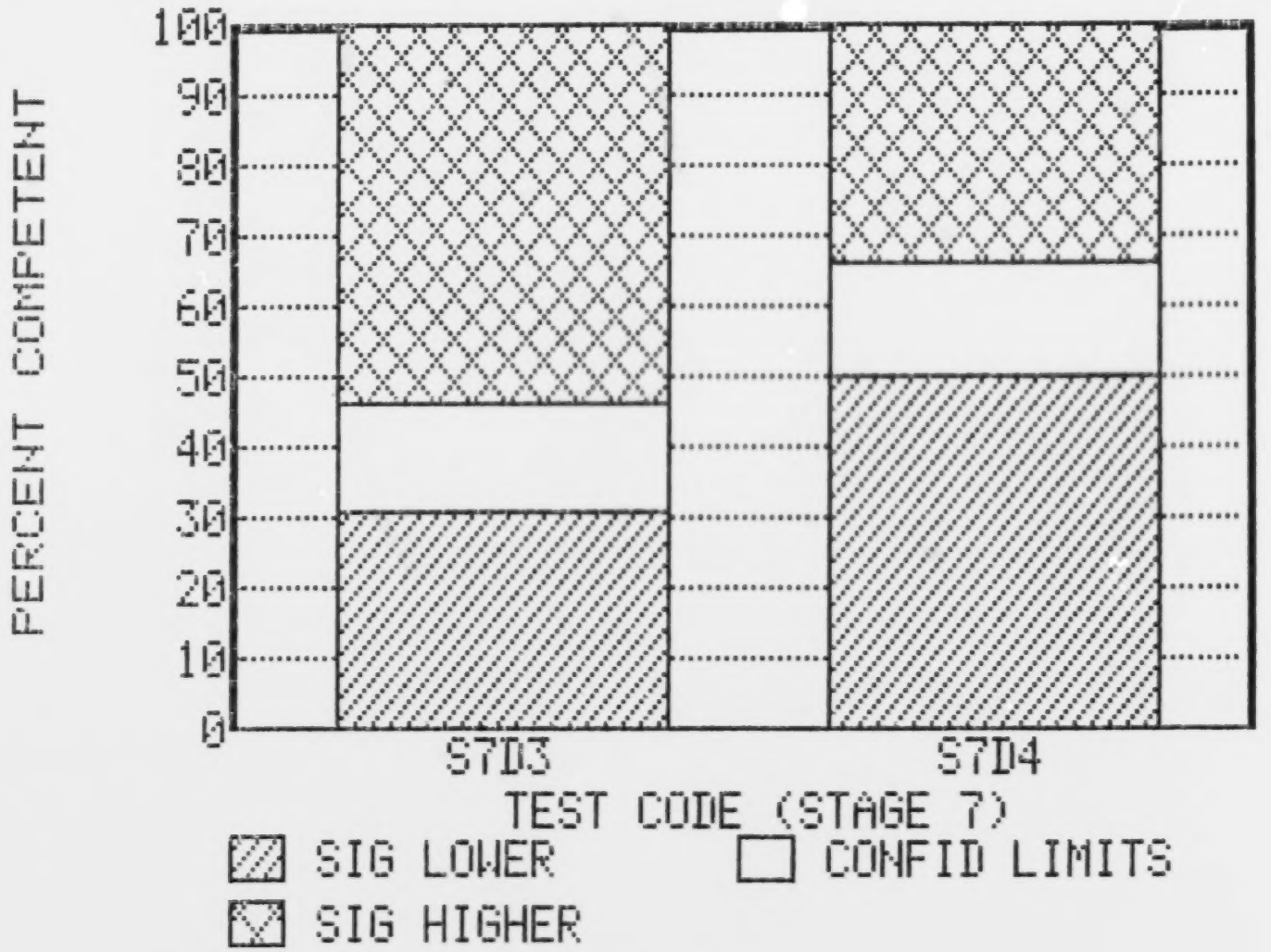


FIG 3.5 DICTIONARY SKILLS



196

197

FIG 3.6 RDP

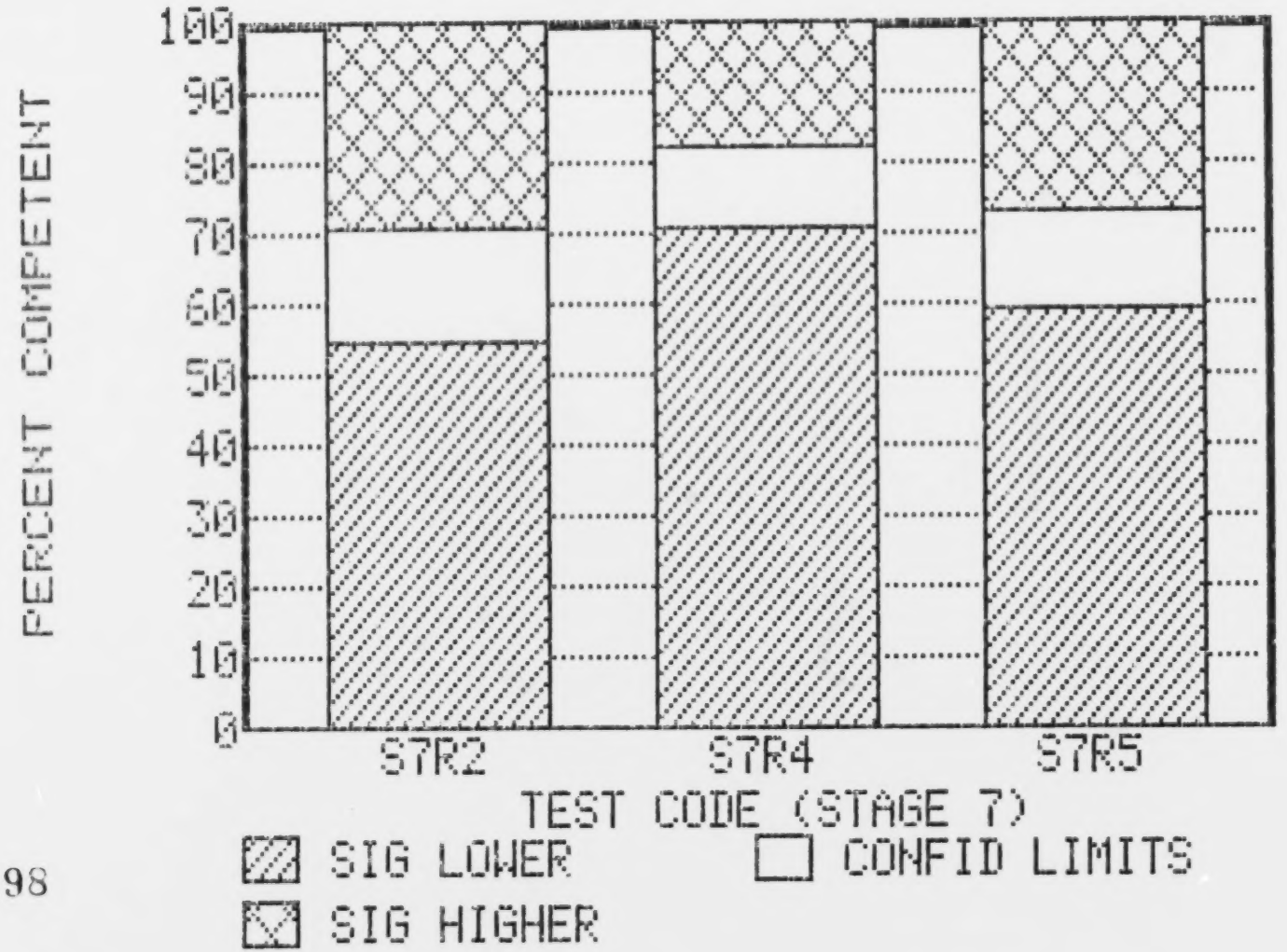

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S7R2
S7R4
S7R5
TEST CODE (STAGE 7)
SIG LOWER
CONFID LIMITS
SIG HIGHER

FIG 3.7 RDP

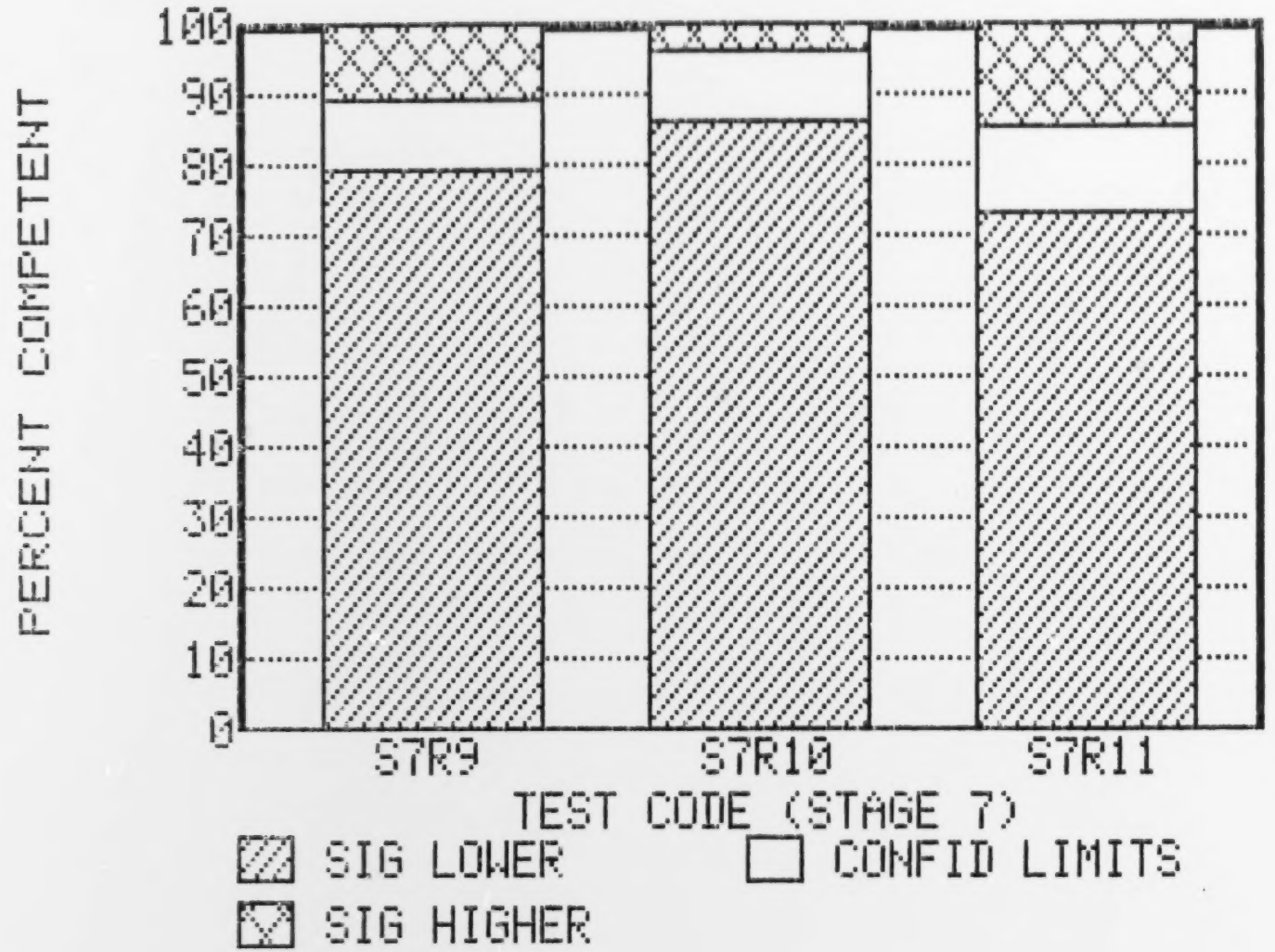

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S7R9
S7R10
S7R11
TEST CODE (STAGE 7)
SIG LOWER
SIG HIGHER
CONFID LIMITS

FIG 3.8 RDP

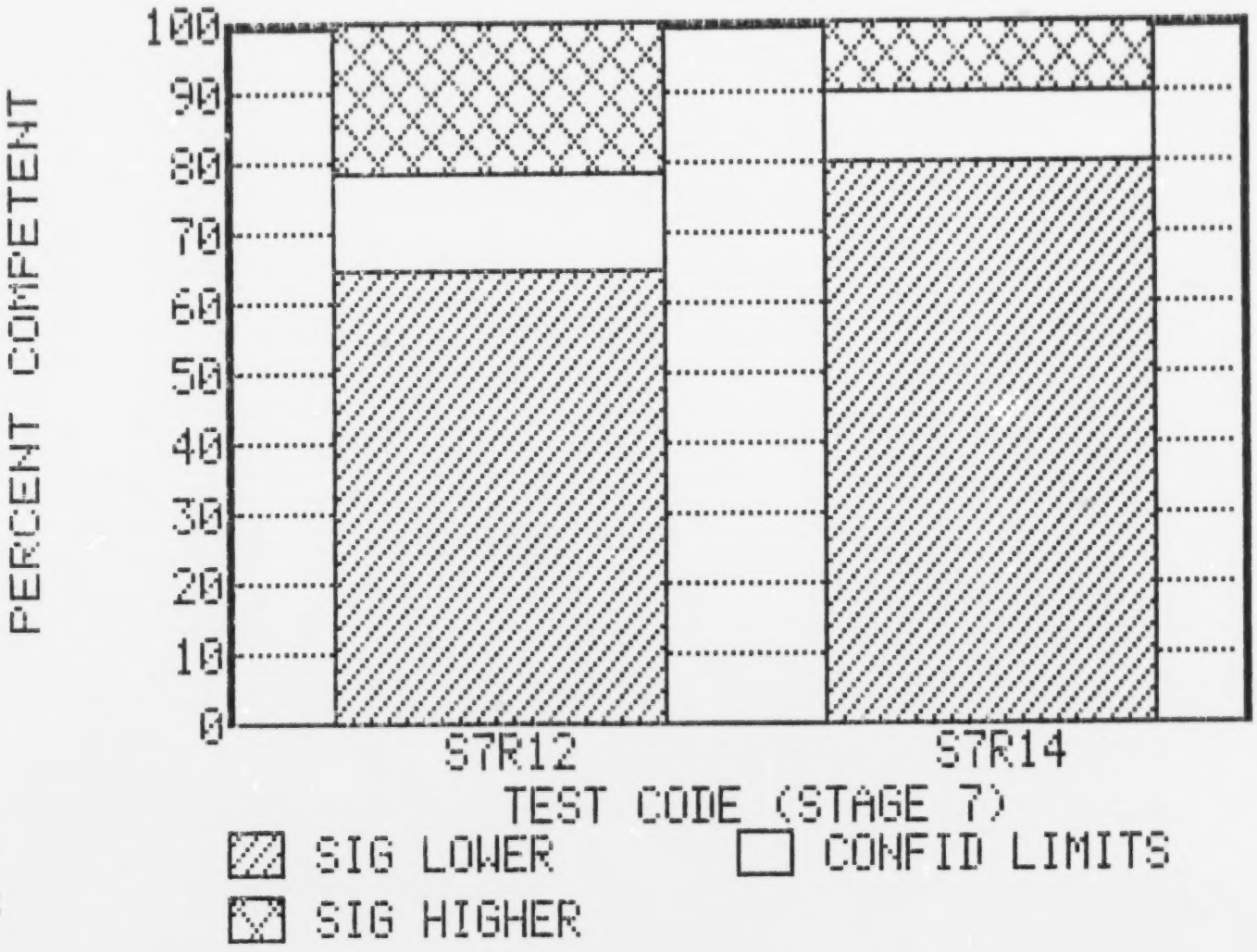

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S7R12
S7R14
TEST CODE (STAGE 7)
SIG LOWER
CONFID LIMITS
SIG HIGHER

FIG 4.1 COMPREHENSION

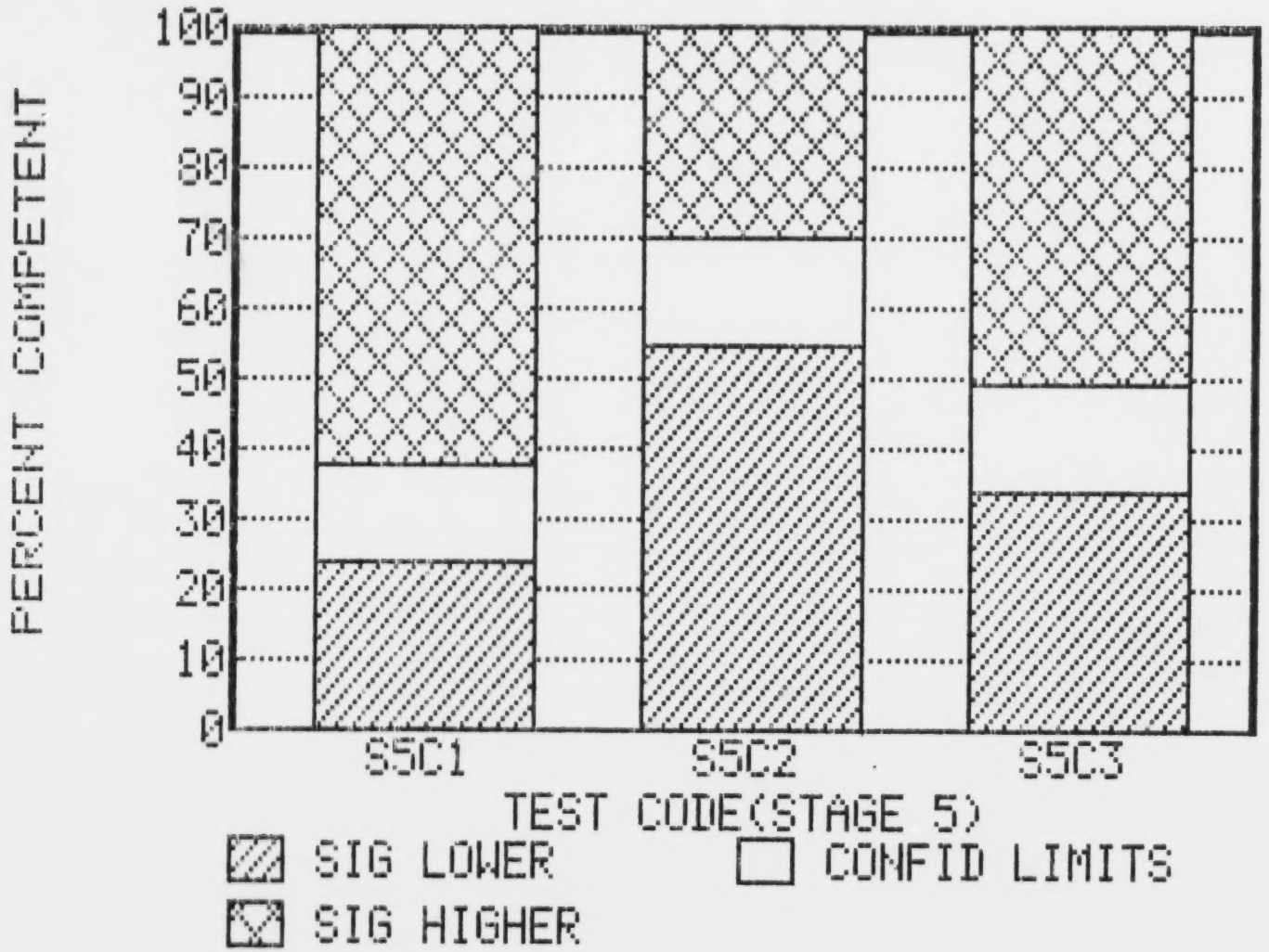


FIG 4.2 COMPREHENSION

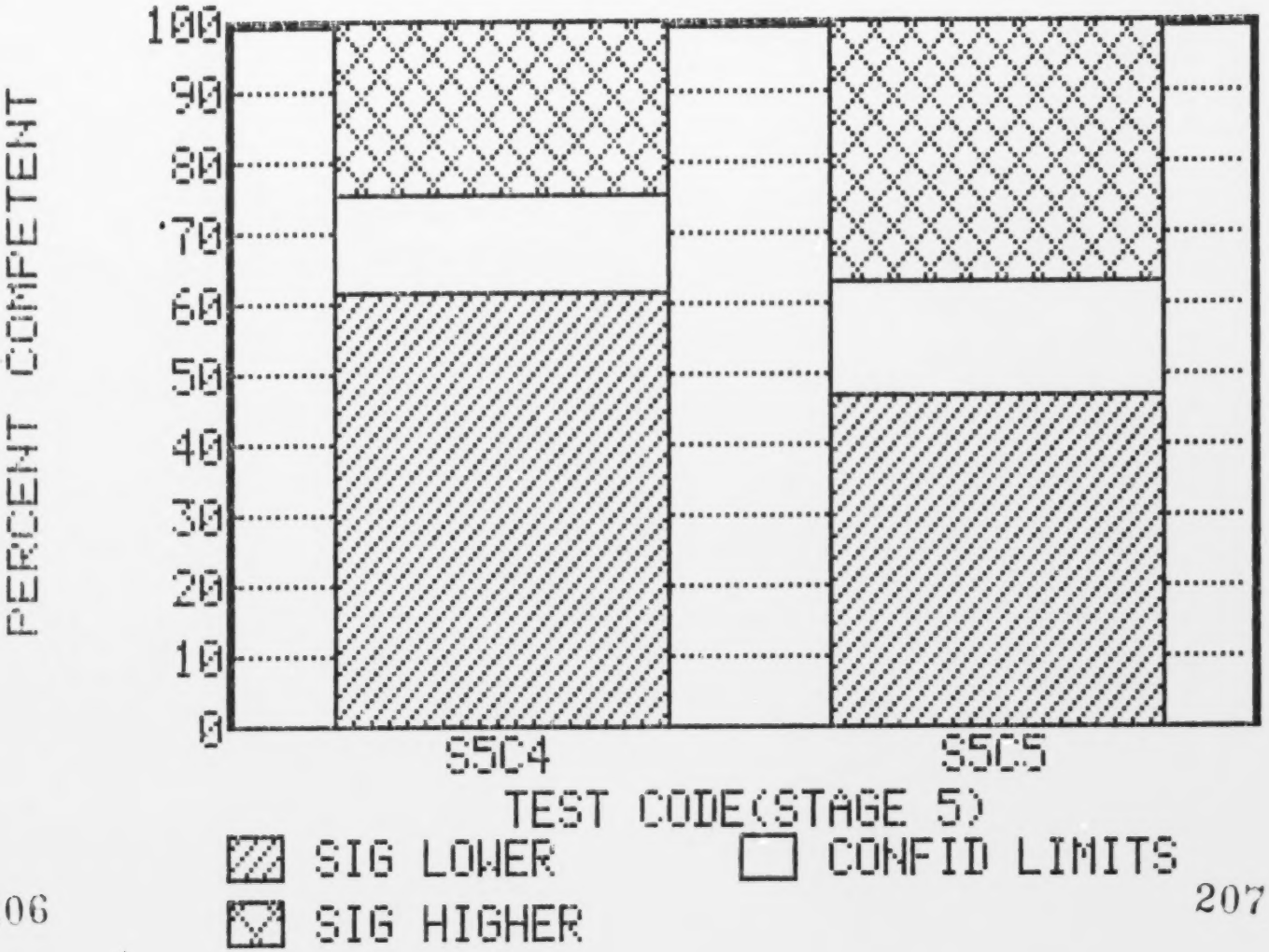


FIG 4.3 COMPREHENSION

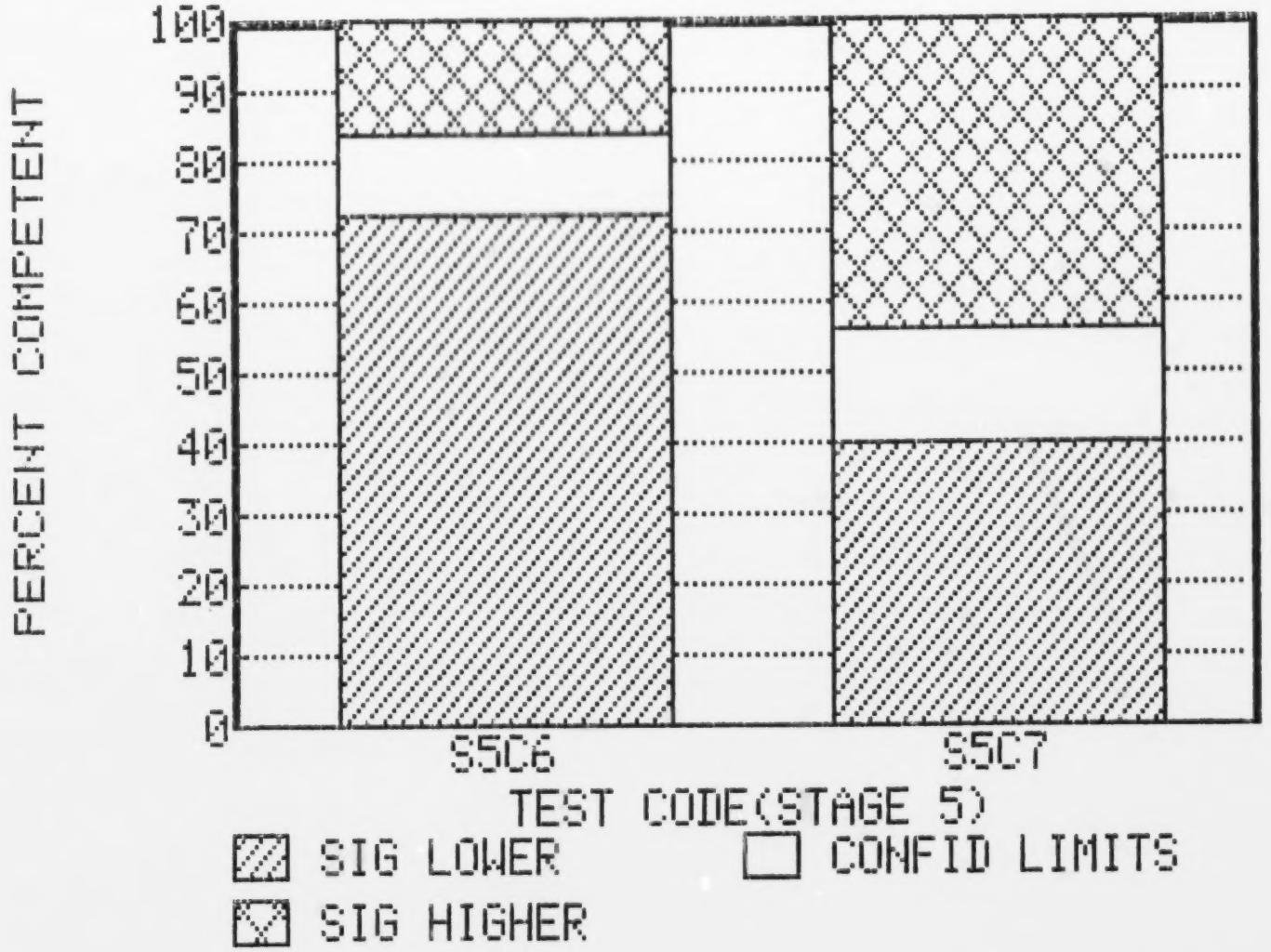


FIG 4.4 DICTIONARY SKILLS

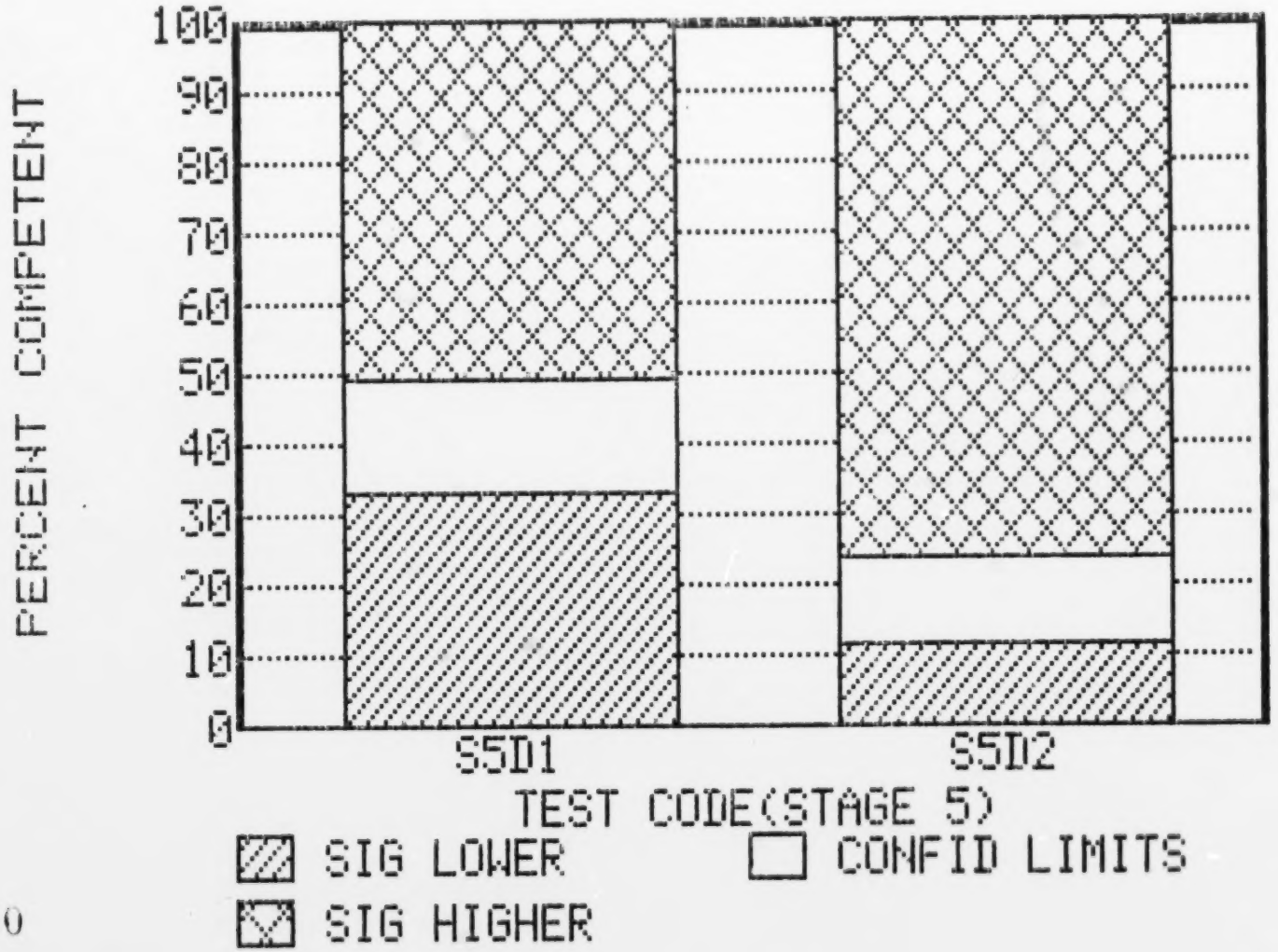


FIG 4.5 RDP

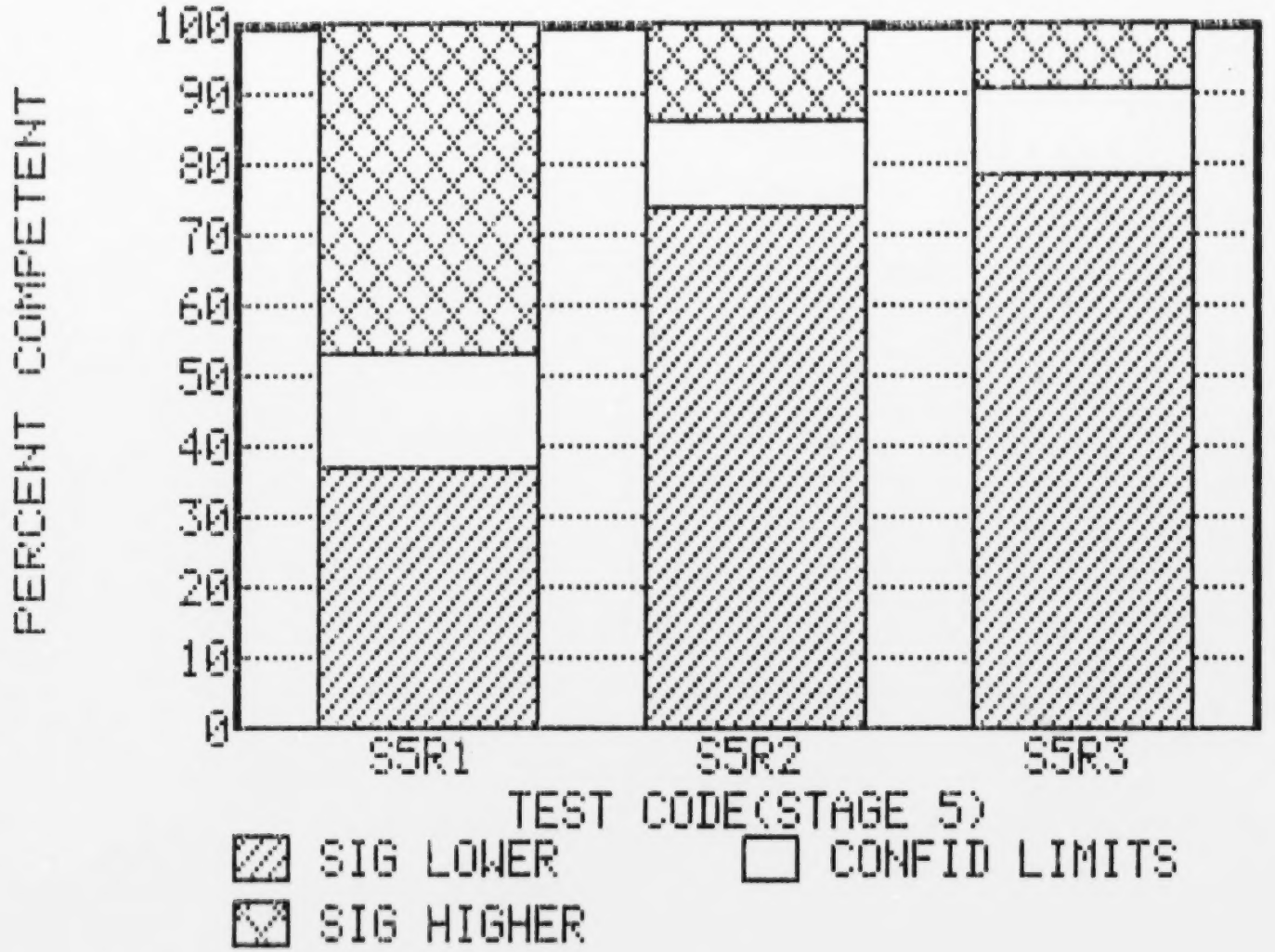

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S5R1
S5R2
S5R3
TEST CODE(STAGE 5)
SIG LOWER
CONFID LIMITS
SIG HIGHER

FIG 4.6 RDP

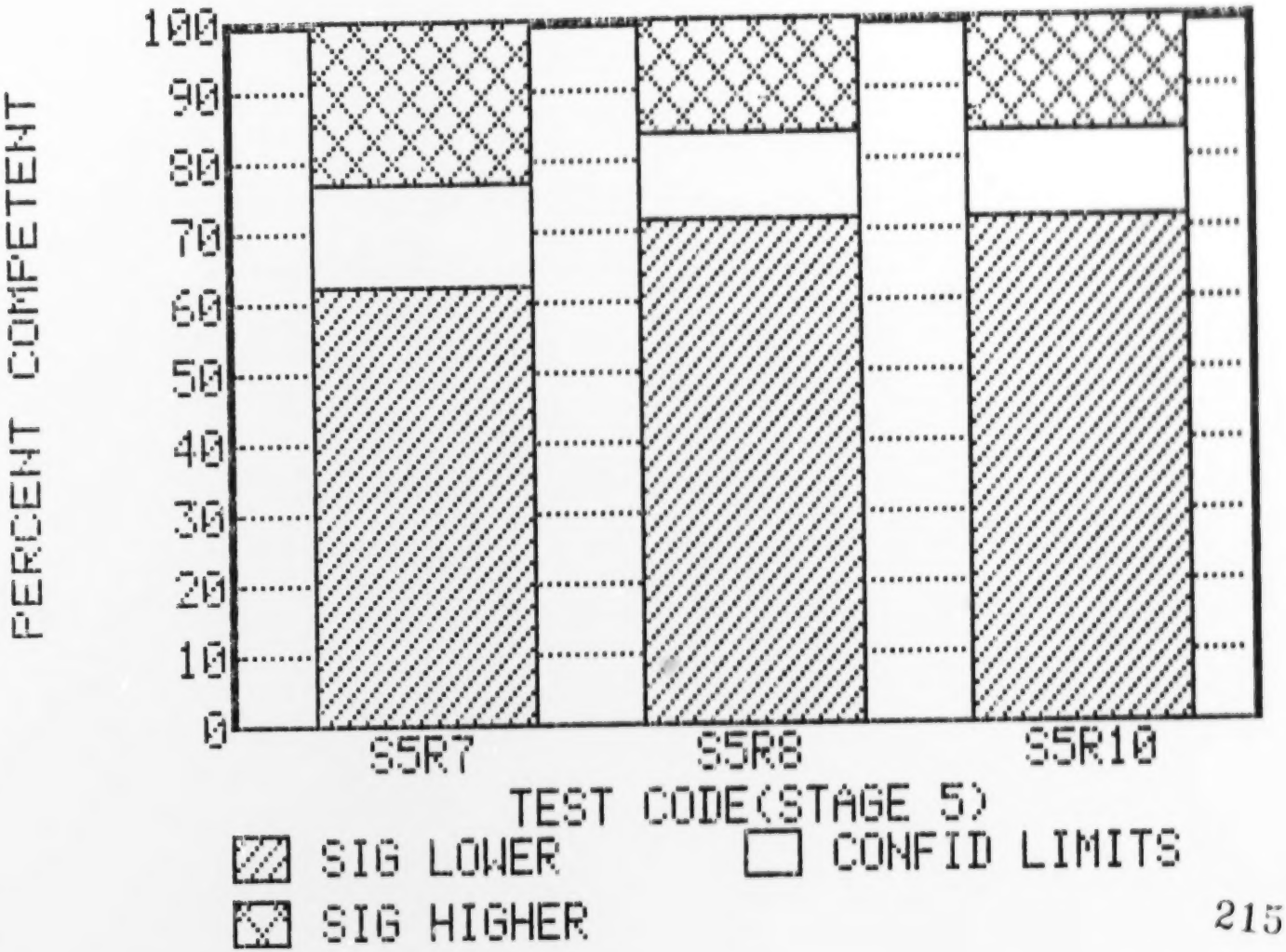
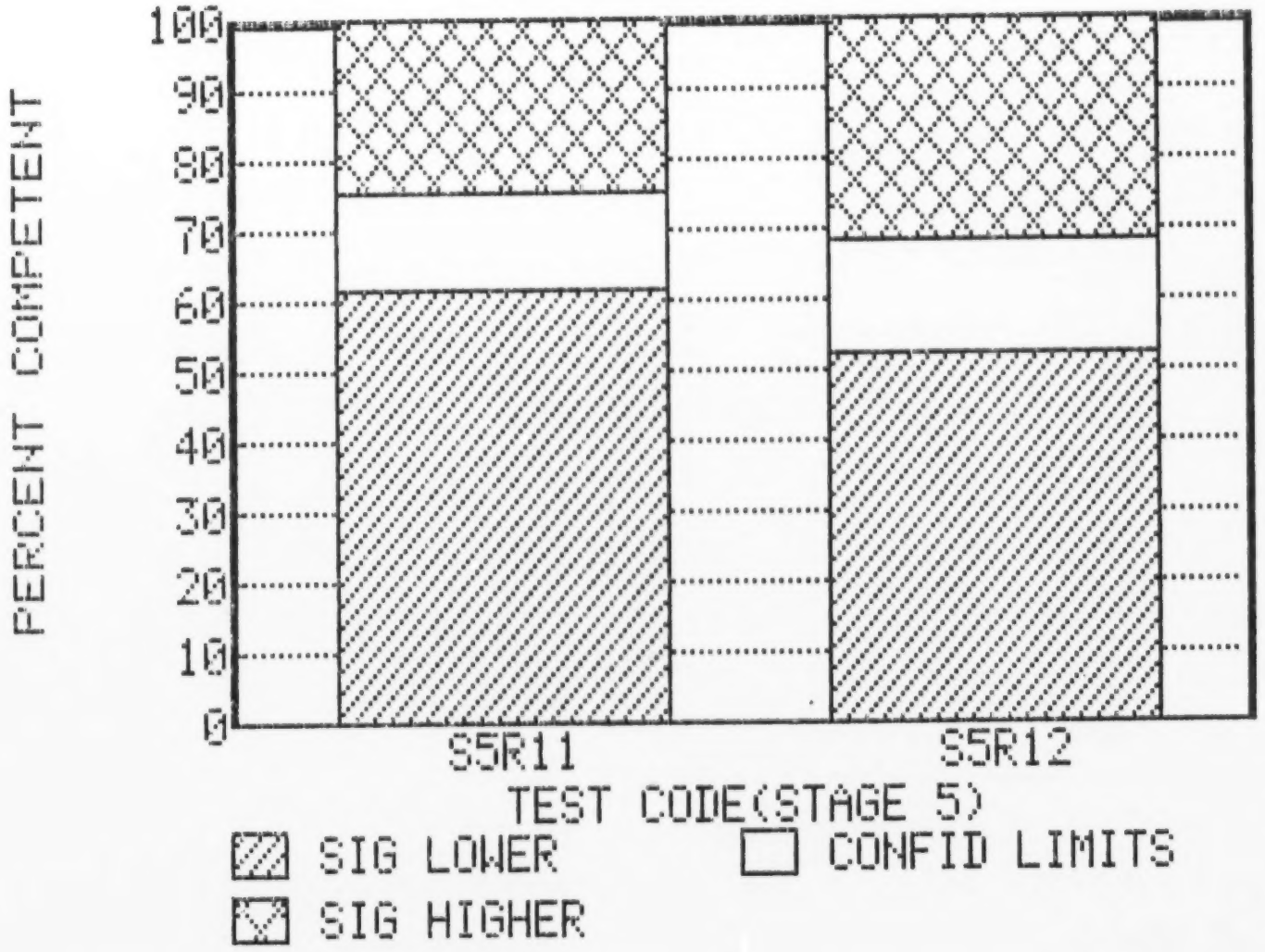

PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S5R7
S5R8
S5R10
TEST CODE(STAGE 5)
SIG LOWER
CONFID LIMITS
SIG HIGHER

FIG 4.7 RDP


PERCENT COMPETENT
100
90
80
70
60
50
40
30
20
10
0
S5R11
S5R12
TEST CODE(STAGE 5)
SIG LOWER
CONFID LIMITS
SIG HIGHER

REFERENCES

Airasian, Peter W and Madaus, George F. Criterion-referenced Testing in the Classroom. **NCME Measurement in Education 3** (1972): 1-8.

Hambleton, R K and Novick, M R. Toward an integration of theory and method for criterion-referenced tests. **Journal of Educational Measurement** (1973).

Hambleton, R K, Swaminathan, H, Algina, J and Coulson, D B. Criterion-referenced tests and measurement: a review of technical issues and developments. **Review of Educational Research** 48 (1978): 1-47.

Izard, J F. **Construction and Analysis of Classroom Tests.** Australian Council for Educational Research (1977).

Kish, L. **Survey Sampling.** New York: John Wiley (1965).

Linn, R L. Issues of reliability in measurement for competency-based programs. In M A Bunda and J R Sanders (eds) **Practices and Problems in Competency-based Education** Washington DC: National Council on Measurement in Education (1979).

Pollock, J E. **Student Assessment: An Inservice Paper for Teachers in NT Schools.** Northern Territory Department of Education (1981).

Popham, W J. **Criterion-Referenced Measurement.** Englewood Cliffs, N J : Prentice-Hall (1978).

Ross, K N. **Searching for Uncertainty,** Melbourne: Australian Council for Educational Research (1976).

Smith, J. **Assessing Developed Abilities in the Northern Territory.** Paper presented at the 10th Conference, International Association for Educational Assessment, Perth (1984).

Wee, C S. **Australian Studies in Student Performance Project 1980: Performance of Ten Year Old Students in Reading and Numeration in the Northern Territory.** Northern Territory Department of Education (1982).

Wee, C S. **Testing for Student Competency in NT Schools.** Paper presented at the Fifth State Research Officers Meeting, Perth (1984).

Wee, C S. **Report on the Performance of Students in the Assessment Program for Urban Primary Schools in the Northern Territory: 1984.** Northern Territory Department of Education (1985).

Wee, C S. **Report on Student Performance in the Northern Territory Primary Assessment Program for Urban Schools: 1985.** Northern Territory Department of Education (1986).